# Review of Gavi's Performance-Based Funding (PBF) Component of its Health System Strengthening (HSS) Support

**September 13, 2018**

**Commissioned by:**
Gavi, the Vaccine Alliance
Global Health Campus
Chemin du Pommier 40
1218 Grand-Saconnex
Geneva
Switzerland


**Prepared by:**
**The Center for Policy Impact in Global Health**
**Contact: Gavin Yamey**
Duke Global Health Institute
310 Trent Drive, Rm 113 Trent Hall
Duke Box 90519, Durham, NC 27708

**In collaboration with:**
**Open Consultants**
**Contact: Marco Schäferhoff**
Cantianstr. 22
10437 Berlin
Germany





**Authors:** Marco Schäferhoff, Gavin Yamey, Osondu Ogbuoji, Yolande Pokam Tchuisseu

## Table of Contents

Annexes are provided in a separate document.

**Tables, figures, and boxes**

## Abbreviations

| | |
|---|---|
| A&R Team | Application & Review Team |
| APR | Annual Progress Report |
| cMYP | comprehensive Multi-Year Plan |
| CSO | Civil Society Organisation |
| DQA | Data quality audit |
| DTP3 | Diphtheria-tetanus-pertussis vaccine, third dose |
| PBF-TT | PBF Task Team |
| EPI | Expanded Programme on Immunization |
| Gavi | Gavi, the Vaccine Alliance |
| GPOBA | Global Programme on Output Based Aid |
| HLRP | High-Level Review Panel |
| HRITF | Health Results Innovation Trust Fund |
| HSCC | Health Sector Coordination Committee |
| HSS | Health System Strengthening |
| IBD | Inter-American Development Bank |
| IHME | Institute for Health Metrics and Evaluation |
| IRC | Independent Review Committee |
| ISS | Immunisation Services Support |
| JRF | Joint Reporting Form |
| MCA | Millennium Challenge Account |
| MCH | Maternal and child health |
| MCV1 | Measles-containing-vaccine first-dose |
| MoH | Ministry of Health |
| M&E | Monitoring and Evaluation |
| NVS | New and underused Vaccine Support |
| PBF | Performance-Based Funding |
| PFA | Partnership Framework Agreement |
| P4P | Pay-for-performance |
| PO | Program Officer |
| RCT | Randomized controlled trial |
| RBF | Results-based financing |
| SCM | Senior Country Manager |
| SMI | Salud Mesoamérica Initiative |
| ToC | Theory of change |
| WHO | World Health Organization |
| WUENIC | WHO/UNICEF Estimates of National Immunization Coverage |
| UNICEF | United Nations Children's Fund |

## Executive Summary

### Objectives and methods

This review evaluates the design, implementation, and results of the Performance Based Funding (PBF) component of the Health System Strengthening (HSS) support of Gavi, the Vaccine Alliance (Gavi). It was commissioned by Gavi and has two key objectives:

- **Objective 1:** To assess the design, implementation, and results of the PBF component of Gavi's HSS support at both global and country levels.
- **Objective 2:** To provide actionable recommendations to inform the future of the PBF component of Gavi's HSS support.

The review of Gavi's PBF uses an evaluation framework based on four dimensions: (i) PBF design, (ii) implementation, (iii) results (ability to earn reward payments, motivational aspects of the reward, changes in coverage and equity), and (iv) lessons learned. The evaluation was based on **key informant interviews** with the Gavi secretariat (26 interviews), former Gavi staff and Gavi's global partners (17 interviews), and country level stakeholders (31 interviews); **review of relevant written materials** (e.g. documents from the Gavi Board and PPC); **analysis of financial data** on eligibility, performance payments, and disbursements; a **benchmarking exercise** to compare Gavi's PBF approach with that of the Global Fund and the Mesoamerican Health Initiative; **review of the key literature on PBF and results-based financing (RBF)** for health improvement in low- and middle-income countries to establish criteria for assessing the design of Gavi's PBF; and five **country deep dives** (Burundi, Ethiopia, Lao PDR, Sudan, and Tanzania).

### PBF design

**Gavi's PBF model suffers from several design weaknesses that are greatly hindering its success**. A major design flaw is that countries with weak immunization data find it extremely difficult to ever qualify for a performance payment.[1] Current measurement systems are not precise enough to measure the small changes in coverage that are the basis for the reward. In addition, achieving increases in immunization coverage within a one-year timeframe is very difficult, and outcome indicators are also sensitive to a range of co-founding factors. Furthermore, while the PBF model helped to boost immunization efforts in some countries, others consider the incentive as too small to have a major motivational effect.[2] There is also a long delay between the successful implementation that triggers a reward, and the actual receipt of the reward. On a positive note, country flexibility in how the reward is used is a valuable design feature.

[1] This is particularly true for countries with DTP3 coverage below 90% at baseline, as these countries are rewarded for (often small) increases in coverage rates rather than for maintaining DTP3 coverage above 90%.
[2] Key informants reported that other Gavi support and other donor funding is much more substantial compared to the PBF amounts. We modeled the potential award for countries with DTP coverage of <90%. Overall, we found that the financial incentive can be substantial but that it differs across countries, depending on coverage levels and the size of the birth cohort.

**PBF should always be accompanied by evaluation and learning—yet we found that Gavi did not embrace the originally envisioned learning agenda.** Given that PBF as a model remains "experimental" (systematic literature reviews show that its effectiveness remains unclear), its use should always be accompanied by evaluation and learning. Indeed, 2011 Gavi Board documents highlighted the need to closely monitor the results of the PBF model, and Gavi's own briefing note on PBF states: "given that GAVI's PBF approach is new, learning from the first phase of countries will be applied to improve the PBF approach in the future." However, we found that Gavi's PBF was not embedded in a learning environment. There is very little evidence on and monitoring of PBF both at global and country levels, which makes it difficult to measure the effectiveness of the PBF model and impedes learning. There is very little specific information on PBF in country proposals and joint reporting forms (JRFs). Results are usually merged together with the overall progress data of Gavi's HSS support. Similarly, the financial reporting usually does not separate between general activities under the HSS grant and those funded by performance payments. There are no PBF reporting templates, and there is no guidance on the frequency and type of reporting for PBF. Thus, Gavi never laid out the conditions for learning because the tools for monitoring the results of PBF were never developed. There is no strategic discussion of PBF at the Board or in the PPC, with reporting currently only taking place at PPC level (and at PPC level, high-level PBF financial data are only provided in annexes).

## Implementation

**There have been problems with communication about the PBF model. In many countries, there is poor understanding of Gavi's PBF, which impedes effective implementation of the PBF model.** While clear communication about the PBF model is a critical precondition to ensure its effective implementation, some countries do not know that the model even exists, let alone how to qualify for the reward payment. In particular, countries that never qualified for a performance payment have very little knowledge of the PBF scheme. Countries will not be motivated to focus more strongly on immunization if they are not aware of the basic elements of the PBF model.

**The alignment of Gavi's PBF with country processes is limited and the PBF model incurs additional transaction costs at global and country level.** The verification and approval process results in a lack of alignment with country planning processes. However, compared with other HSS grant processes, countries find the PBF approach to be less burdensome. Gavi's PBF is also not always fully integrated with other RBF schemes at country level, a missed opportunity for alignment and harmonization.

**Countries have mostly used the PBF funding to fill gaps in their budgets.** These gaps have included salaries, bonuses, supervision, warehouses, vehicles, and surveillance. This wide range of uses of the performance payment reflects the fact that Gavi has not given countries sufficient guidance or direction when it comes to the PBF model. Gavi's original idea was to be "light touch" and deliberately allow flexibility in how performance payments are used, in the hope of sparking creativity and innovation in immunization financing. However, the result of this approach is that countries are generally using the money to do "more of the same." Many countries, including Tanzania, Lao PDR, and Mozambique, also use the PBF reward to co-finance the Gavi's Cold Chain

Equipment Optimisation Platform, which is helping countries modernize cold chains with high-performing equipment.

**Countries that are already struggling to use their HSS grants because of poor absorptive capacity can still qualify for PBF payments—not surprisingly, they typically also struggle to effectively use the reward.** Progress in the implementation of the HSS grant does not affect the decision about PBF eligibility. For example, countries can have low use of their HSS grants but can still earn additional funds through the PBF model. Countries may struggle to effectively use the additional funds in light of already existing absorption issues—a challenge that cuts across countries and that is unrelated to population size. Thus, the PBF payments become less relevant in such countries.

## Results

***Lack of equitable access: Distribution of performance payments across countries*** **Gavi's PBF model mostly benefits countries with high (≥90% DTP3) baseline coverage. Countries with low baseline coverage (<90% DTP3) find it difficult to earn reward payments due to data issues and/or lack of progress.** Over the period 2014-2017, 31 distinct countries were assessed for eligibility (some countries were assessed multiple times over multiple years—there were 69 assessments in total). A total of 16 different countries earned a reward in at least one of these four years.[3] The total amount of performance payments was US$34.8 million over the 2014-2017 period. A striking finding of our analysis is that of the 16 countries that received a reward payment, 11 had a baseline DTP3 coverage of at least 90% (high coverage countries) and only five had coverage below 90% (low coverage countries). The success rate for receiving a reward payment was *much higher* for the high coverage countries than the low coverage countries: 11 out of the 12 high coverage countries that were assessed for eligibility (92%) received the reward compared with just five out of 19 low coverage countries assessed for eligibility (26%).

Looking at the payment amounts, the **11 countries with high baseline coverage accounted for 87.5% of total allocated payments** (US$30.4 million), while the five countries with baseline coverage below 90% only earned 12.5% of total payments (US$4.3 million). A similar pattern was seen with disbursements: high baseline coverage countries received 80% of total PBF disbursements (US$12.5 million) and low baseline coverage countries only 20% (US$3.1 million).

For the period 2014-2017, high baseline coverage countries were **6.9 times more likely** to be eligible for a reward based on DTP3 performance, and **4.5 times more likely** to be eligible for *any* performance payment (i.e. DTP3 and/or MCV or geographic equity performance payments).

DTP3 payments to low baseline coverage countries continued to decline throughout the period with **no country from the low baseline coverage group receiving a performance payment for increased DTP3 coverage in 2017**.

**The verification mechanism has prevented countries with weak data from accessing performance rewards.** In 2017, 50% of cases of ineligibility were due to poor data quality, 43% due to

[3] Ten out of these 16 countries received awards in multiple years (Burundi, Lao PDR, Nicaragua, Rwanda, Sudan, Tanzania, Zimbabwe, Honduras, Korea, and Solomon Islands).

stagnating/decreasing coverage, and 7% to both. Often, this ineligibility results from the substantial differences between the data reported by countries and the WUENIC data. The countries that are least likely to pass the verification process are often those with the weakest health systems and thus the greatest need.

**Francophone and fragile countries rarely qualify for performance payments.** Only two francophone countries from Africa ever qualified for the performance payment, and only 4 out of 18 countries from Gavi's fragility list ever did.[4]

***Motivational effect***

**Gavi's PBF model is valued by countries that qualified for performance payments in the past.** Countries that have qualified for the performance payment gave the PBF model high ratings for relevance and overall success. The average score across countries was 3.5 for relevance (using a scoring scale of 1-4, where 4 is highest relevance) and 7.1 for success (using a scoring scale of 1-10, where 10 is the highest level of success).

**Our study found mixed results on whether the PBF had a motivational effect.** As part of the key informant interviews, which were qualitative in nature (although two interview questions asked key informants to give a numerical "rating"), we asked countries whether the PBF model incentivized them to improve their immunization programs. Most countries qualifying for performance payments between 2014-2017 reported that the PBF incentivized better immunization planning and implementation and helped them to focus on the equity agenda. For example, in Nicaragua the core funding from the HSS grant was programmed so that low-coverage districts were further prioritized. In Sudan, the reward payments encouraged efforts to reach the hardest to reach populations (e.g., conflict-affected areas, or low-density areas with nomadic populations). Other countries reported that the focus on equity stimulated discussions at the subnational level on ways to ensure that no child was left out. In turn, these discussions led to improved planning and prioritization and reallocation of resources (for example, through revisions of the countries' RED strategy). In addition to the development and revision of national and subnational plans, countries reported that the PBF contributed to improved guidelines and policies. Some countries also worked on the reporting processes at district level to make the qualification for future performance payments more likely. Country stakeholders also reported that Gavi's PBF helped to introduce a performance-based service delivery culture, which shifted the focus from inputs towards measuring outputs – a fact that subsequently also helped other funders with their performance-based programs.

There is clearly an opportunity for Gavi to learn lessons from countries that reported a positive motivational effect from the incentive payment. Clearly, there are countries that mentioned during the interviews that the PBF model did not incentivize them. These countries gave different reasons why they were not incentivized, including challenges related to data and verification, problems in the implementation of the general HSS grant, or broader issues like political instability and unrest.

[4] At the time of writing this report, 18 countries were on Gavi's fragility list; since then, Nigeria and Ethiopia have been removed.

**Several countries that received a PBF payment showed rising immunization coverage and improvements in equity but a causal link between the payment and changes in coverage/equity cannot be proved.** Given that there has been no formal impact evaluation using randomization and controls, it is not possible to specifically attribute these changes specifically to the PBF itself. There are no objective, verifiable data linking coverage and equity with the performance payments themselves.

## Lessons learned

**Gavi should keep its PBF model if it ensures learning and is willing to take a risk. If not, the PBF model should be ended.** Overall, our review suggests that Gavi should continue with a PBF model. Countries that qualified for performance payments in the past value the PBF model (these countries gave high ratings for relevance and overall success). Most of these countries also reported that the PBF motivated them to improve their immunization programs. At the same time, there were countries that did not benefit from the PBF due to issues related to its design and because of implementation challenges. It will be important to change the PBF to also make it work for countries with lowerDTP3 coverage (below 90%). We therefore recommend that Gavi redesigns its PBF model by better recognizing the different country contexts rather than ending the PBF model. As described in our recommendations below, we suggest a tiered approach in which the model differs for countries with high baseline coverage ($\geq$90%), medium baseline coverage (70-89%), and low baseline coverage (<70%). In addition, Gavi has been at the forefront of PBF and the interest in PBF among donors and countries is growing, and we are still in a highly active learning phase. It is critical for Gavi to be in the PBF "learning arena."

However, Gavi should only keep the PBF model under two conditions:

- The first is that it ensures learning and provides the conditions for learning within the Secretariat and the whole alliance. The Secretariat needs to develop the necessary tools to effectively monitor the implementation and results of the PBF. Based on the improved monitoring, it should regularly report to the Board and the PPC to allow for discussion and learning.
- Second, Gavi should decide if it is willing to take risk. We believe that Gavi's continued testing of the PBF model will require the organization to be comfortable in accepting a certain level of risk. All PBF models run the risk of creating perverse incentives—in particular, the risk of over-reporting (inflating results in order to receive reward payments). To remove this risk almost entirely requires costly, highly intensive, external verification systems that can end up constituting a huge proportion of the total costs of the PBF scheme. We believe that Gavi's approach of investing in national data systems, rather than external verification, is much more valuable over the long run. Our analysis did not find evidence that countries that had received Health Information System (HIS) support from Gavi were more likely to receive a PBF reward payment but this analysis faced limitations in terms of data. While we thus cannot prove that Gavi's HIS support *led* to higher success, we think that these payments will pay off and contribute to better performance measurement.

**The PBF design needs to be changed. It suffers from being too ambitious and – as discussed above – its underlying assumptions are flawed**. Gavi's PBF has not worked in all countries. The outcome focus is too difficult for many countries (those with baseline DTP3 coverage of <90%), more time is needed to achieve the outcomes, and the reward should be paid more quickly.

During the redesign process, the current model should continue in its current form. Abruptly stopping the model would likely foster confusion, disruption, and unpredictability at country level.

**The PBF design is too sensitive to measurement errors.** The verification is not working well; there are major problems with the data. We ran three different hypothetical models to see which approaches could help to alleviate these data and verification issues:

- First, we assessed whether a "higher tolerance level" (allowing a 10% rather than 5% difference between WUENIC [WHO/UNICEF Estimates of National Immunization Coverage] data and administrative data) would have made a difference in the 2014-17 timeframe. It did not make much difference (country data differs substantially from WUENIC data for non-qualifying countries).
- Second, we compared each country's actual payment experience with the payment experience under a hypothetical WUENIC option (in which countries use WUENIC data for their baselines and the annual PBF reviews for the period 2014-2017). The results show that under the WUENIC option, 17 countries (representing 31 PBF awards) would have had a different payment experience. In this scenario, 15 of 17 countries would have received at least one additional PBF award (of these, 13 had low DTP3 coverage at baseline, two had high coverage), one country would have lost one PBF award, and one country would have lost one DTP3 award but gained one equity award. While the new WUENIC model may thus help to some extent, more reforms will be needed.
- Third, we modeled a scenario where countries with low baseline DTP3 coverage qualified for an equity bonus if they showed an increase in the percent of districts with DTP3 coverage ≥80%, regardless of their performance on other metrics (we note that this is an alternative equity measure than the one currently used by Gavi for high baseline coverage countries). In this scenario, 12 out of the 19 low baseline coverage countries would have qualified for at least one equity PBF bonus while 7 countries would not have qualified (including Ethiopia, which would have qualified for three equity bonuses). This modeled scenario shows the potential to introduce equity indicators to the low baseline coverage group. In fact, the 2011 Board paper on PBF considered introducing an equity indicator to countries with baseline coverage below 90% if the equity indicator proved to be a good measure. However, as there was little learning, this option was not taken further.

## Recommendations

If a learning environment is established and Gavi is willing to take a risk, the PBF model should be continued but with important changes to its design, with regards to implementation, and learning.

***Design recommendations***

**Recommendation 1: The PBF model needs to move away from its current approach, since it does not work for all countries. We recommend that the model should have three country groups: a high coverage group (DTP3 coverage of at least 90% at baseline), a medium coverage group (70-89% DTP3 coverage at baseline), and a low coverage group (under 70% at baseline). The current PBF model works well for the high coverage group. For the medium coverage (70-89%) group, we propose new indicators and decision rules for receiving the reward. For the low coverage group (below 70%), PBF should not be used—instead, countries need to first build their health systems capacity, including health information systems.**

▪ *For the high coverage group (≥90% DTP3 coverage at baseline)*: The current incentive mechanism rewards countries with high coverage to maintain coverage levels; the existing model is good enough for these countries and we believe it can be kept.

▪ *For the medium coverage group (70-89% DTP3 coverage at baseline)*: For this group, we propose a new approach involving three new indicators: (a) a standardized systems indicator that cuts across all countries from this group (measuring stock-outs), (b) a second, country-specific immunization process/systems indicator chosen from a country's grant performance indicators, which reflect Gavi's investments in HSS in individual countries, and (c) an equity indicator. Country performance for these three indicators should not be measured as "pass or fail"—instead progress in these indicators should be measured continuously:

- **Standardized systems indicator:** Intermediate system and/or process indicators rather than coverage indicators should be used to incentivize and reward the medium coverage group. There is strong evidence showing that selected systems/process indicators contribute to coverage and equity. One specific system indicator that could be used for this purpose is *vaccine stock-outs*, as measured, for example, by the proportion of facilities with full availability of all or a selected set of tracer vaccines and immunization supplies over a resupply period. Gavi routinely receives this information as this is a core intermediate indicator from its grant performance framework.[5] Monitoring of stock-outs is routinely conducted by countries as part of the Joint Reporting Process. However, the country-reported data would need to be verified, which will involve additional transaction costs (see Recommendation 3 below).[6]
- **Country-specific system/process indicator**: There should be a second systems or process indicator that would strongly reflect Gavi's HSS investments in individual countries. This second indicator would also be part of the grant performance framework. It could also be a core indicator (like the indicator on stockouts), or a tailored indicator. This country-specific indicator would be selected *by countries themselves* in dialogue with the Gavi secretariat

[5] Gavi, the Vaccine Alliance: Considerations for countries on targeting Gavi investments to achieve immunisation outcomes. Focus Area Immunisation Supply chain. Revised Version May 2018. Available at: https://www.gavi.org/support/process/apply/hss/

[6] Going forward, it would be critical for Gavi to further standardize the way that this stock-out indicator is monitored across countries.

(examples include the proportion of planned immunization outreach sessions conducted; the proportion of health facilities with at least one qualified/trained vaccine provider). Under this second indicator, Gavi could also reward improvements in data accuracy as this will improve the PBF measurement going forward and contribute to improved programming (our suggested verification tool has previously been used to measure such improvements; see Recommendation 3).

- **Equity indicator:** For the medium coverage group, performance payments should also be made conditional on improvements in equity. Gavi could use the equity indicators from the high coverage group, which is in line with the Global Vaccine Action Plan, though we suggest a more radical approach (see Recommendation 2 below).[7] Rather than using this indicator in a "pass or fail" way, continuous progress should be measured.

▪ *For countries with <70% DTP3 coverage at baseline:* PBF should not be used. Instead, Gavi should support these countries to develop their systems first (infrastructure, data systems, etc.) before being eligible for rewards. We arrive at this recommendation based on thorough triangulation of our results. First, the low coverage group comprises 11 countries from the list of countries that are eligible to apply for Gavi support in 2018.[8] Nine of these 11 countries are on Gavi's 2018 list of fragile states, so rather than being low-resource settings only, these countries face an additional set of major systemic challenges. The conditions in these countries are very tough, with large-scale conflict, high risk of disease outbreaks, and very weak or non-committed state actors. Findings from our literature review show that PBF is difficult to implement in fragile states. Multiple studies have concluded that country fragility is a major barrier to the success of PBF due to problems such as weak health information systems, governance, and leadership. Second, Gavi's own PBF data shows that the implementation of PBF models is difficult for the low coverage group. No country with less than 70% DTP3 coverage at baseline qualified for a performance payment between 2014 and 2017. Only four fragile countries qualified for a PBF performance payment in this period. Gavi also already recognized the challenges associated with these fragile settings, and provided the performance payments to fragile countries like Afghanistan and Somalia under the country tailored approach, although they did not qualify for these payments. Third, key experts that were interviewed also emphasized the difficulties in fragile countries, supporting our approach to the low coverage group. We thus believe that Gavi should not use PBF for the low coverage group. Once the countries are ready, and have been accredited, they can be enrolled into the PBF scheme.

The recommended segmentation involves different mechanisms for each group by which the incentive could link to performance. High coverage countries would continue to be rewarded for maintaining coverage at a minimum of 90%, while countries from the medium coverage group need to show improvements on key system indicators. However, maintaining coverage at a high level (at least 90%) signals strong continuous performance across the system and countries should be rewarded for this achievement. For the medium coverage group, a focus on system indicators is much more appropriate and – together with an improved verification approach - should lead to more

[7] http://www.who.int/immunization/global_vaccine_action_plan/GVAP_doc_2011_2020/en/
[8] https://www.gavi.org/support/sustainability/countries-eligible-for-support/

balanced PBF results, i.e. to a model that works for all countries even if different mechanisms are at play.[9]

**Recommendation 2:** We believe Gavi should consider a more radical measure for equity for both the high coverage and the medium coverage groups. Gavi's goal of increasing *national* coverage is not necessarily in line with its current equity efforts as geographic areas that already lag behind might receive even less attention as they are the hardest to reach.[10] We suggest that the PBF model should be targeted at increasing coverage in three areas of very low coverage ("hot spots"). Periodic population-based surveys provide data that could pinpoint parts of the country that are under-served hot spots. Gavi could incentivize countries to increase their attention to these areas. To measure performance, Gavi would require additional monitoring of performance in these hot spots as a supplement to more standard, nationwide measures of the immunization system and coverage. Gavi could require countries to focus on a subset of those regions/districts with the largest percentage/number of zero-dose (zero-antigen) children who never received a single dose of a vaccine. The performance payment would then be paid based on the additional number of previously zero-dosed children.[11] As an alternative to rewarding additional zero-dosed children immunized in these hot spots, Gavi could reward the additional number of children immunized.

***Implementation recommendations***

**Recommendation 3: On data verification, the current system is working well enough for the high-performing countries, but a new system is needed for the 70-89% coverage group.** For the 70-89% coverage group, there will need to be a new mechanism put in place to verify the reported system indicators, including stock-outs. Gavi should avoid using very costly and labor-intensive verification approaches (e.g., those used by the World Bank or the SMI). One cost-effective approach to verification would be to conduct small sample size surveys, which are cheaper than large sample size surveys. One example of such a survey is lot quality assurance sampling (LQAS), which is relatively simple to conduct and has been used by other major organizations working in global health. These LQAS surveys could also contribute to measuring the overall performance of the larger HSS grant, as the marginal cost for including additional measures (or indicators) to an existing LQAS survey will be minimal. This approach would prevent the creation of a standalone PBF verification mechanism; instead, it would establish a mechanism that feeds into the overall HSS grant. LQAS could also be used to measure the suggested equity indicator, and it has also been used to validate improvements in data accuracy. While LQAS surveys come at a cost, which depends on the size of the survey and country context, they are still the most inexpensive verification mechanism (it is true that the current mechanism comes at zero cost but our review has shown that it is not working). And as highlighted,

[9] For both the high and medium coverage groups, we suggest that there are no changes in the relationship between the programmable payment and the performance payment - the current model ensures sufficient predictability.
[10] In addition to the spatial dimension of equity, there are other dimensions. Groups from the lowest socioeconomic quintile, and/or marginalized and discriminated groups often have less access to health services.
[11] This information is regularly available from standard Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS), so baseline data would be available. DHS and MICS are typically conducted every five years, but are staggered, i.e., data are available about every 2.5 years (thus monitoring for Gavi's PBF would incur additional costs if data are needed more frequently).

LQAS surveys have the potential to become a valuable institutionalized mechanism to independently assess data quality and track changes over time.

**Recommendation 4: Gavi needs to significantly improve its communication to countries and Gavi staff about the PBF mechanism.** Our study very clearly found a need for better communication about what the PBF model is, how a country qualifies for it, and how it operates. Rather than intensifying communications on PBF immediately, we suggest strengthening PBF communications when the revised PBF model is launched across countries.

***Learning recommendation***

**Recommendation 5: There needs to be better reporting of the results of Gavi's PBF and a stronger culture of learning.** The Secretariat should develop tools to effectively monitor the implementation and results of the PBF, which should leverage joint appraisals. In addition, changes to the guidelines for applying for Gavi support might be required. Based on the data generated from these tools, it should report to the Board and the PPC on an annual basis to allow for discussion and learning. Progress will also depend heavily on learning from the best-performing countries in recent years, so Gavi should support South-to-South learning to ensure truly transformative shifts even in the poorest countries with the lowest coverage. This South-to South learning will also help to facilitate a more strategic and innovative use of PBF rewards across countries.

# 1. Introduction

## 1.1 Objectives of the review

This review evaluates the design, implementation and results of the Performance Based Funding (PBF) component of the Health System Strengthening (HSS) support of Gavi, the Vaccine Alliance (Gavi).[12] It has two key objectives:

- **Objective 1:** To assess the design, implementation, and results of the PBF component of Gavi's HSS support at both global and country levels.
- **Objective 2:** To provide actionable recommendations to inform the future of the PBF component of Gavi's HSS support.

As part of Objective 2, the review also provides recommendations on if, and how, Gavi should restructure PBF to be relevant, effective, and efficient to contribute to achieving Gavi's 2016-20 strategy.

The review covers the period starting in 2009, the year when the redesign of the PBF approach began, to December 2017. However, the more specific focus of this review was on two strategic periods, Phase III (2011-2015) and Phase IV (2016-2020).

The review has been commissioned by Gavi, and its primary audience is Gavi, particularly the Gavi Alliance Board (the Board), the Performance & Policy Committee (PPC), and the Gavi Secretariat (the Secretariat).

The review is organized as follows: In the following section (1.2), we introduce the PBF model. Section 2 provides an overview of the methodology of the review. Section 3 establishes key criteria for the assessment of the design of the PBF model, including from a literature review and a comparison with two similar PBF initiatives (the Salud Mesoamerica Initiative [SMI] and the Global Fund's PBF approach). Section 4 presents the findings on the development and the design of Gavi's PBF model. Section 5 focuses on the results of the implementation of the PBF model. Section 7 summarizes the lessons learned. Section 8 makes recommendations for the redesign of Gavi's PBF.

## 1.2 Gavi's PBF model

In a pioneering effort, Gavi introduced its original PBF approach in its first strategic phase (2000-2006) as part of its Immunisation Services Support (ISS). ISS provided results-based funding for the strengthening of immunization systems based on the number of children vaccinated with three doses of diphtheria-tetanus-pertussis (DTP3). ISS was phased out in 2008 and the Gavi Board convened a

[12] Gavi commits to HSS grants up to a five-year period, with the first one or two tranches usually approved with the approval of the proposal. In subsequent years, countries must submit a renewal request in order to request the next HSS funding tranche. The objective of HSS grants is "to sustainably address health system bottlenecks to equity in immunisation coverage"; see Gavi, the Vaccine Alliance 2016: Gavi Alliance Health Systems and Immunisation Strengthening (HSIS) Framework.

PBF Task Team (PBF-TT) in 2009 to identify opportunities for incorporating PBF into Gavi's operating model and to better support countries with low routine immunization coverage.

In November 2011, the Board decided to integrate PBF into Gavi's HSS support by rolling out a PBF component for all HSS grants approved in 2012 and onwards.[13] Following this decision, the HSS funding window included two different types of payments:

(i) a **programmed payment**: yearly tranches for which disbursement would be based on progress in implementation and on achievement of intermediate results; and

(ii) a **performance payment**: yearly instalments for which eligibility and amount would be based on improvements in immunization outcomes.

The main criteria for the receipt of performance payments were set in 2014 and involve different conditions for countries with < 90% DTP3 coverage at baseline and those with ≥ 90% DTP3 coverage at baseline (see Figure 1 and Table 1).[14] In summary, the key components of the PBF model are:

► In year 1, countries can receive an upfront investment of up to 100% of the country ceiling, and then in subsequent years they can budget up to 80% (this is the programmed payment).[15]

► From year 2, countries can earn additional annual performance payments that can reach up to 150% of the country ceiling for countries with < 90% DTP3 coverage at baseline or up to 120% of the country ceiling for countries with ≥ 90% DTP3 coverage at baseline.

► Gavi uses a set of decision rules to determine whether a country receives the performance payment (Table 1). These rules are based on verified coverage of DTP3 and routine measles first dose (MCV1). The coverage produced by the country's administrative data system must not be more than 5% higher than the WHO/UNICEF Estimates of National Immunization Coverage (WUENIC). New measurement mechanisms were introduced in 2015, including the option for countries to to use the WUENIC data (rather than administrative data) for their baselines and the annual PBF reviews, but these are not widely used yet (Box 1).

►In June 2016, the Gavi Board approved the Health System and Immunisation Strengthening (HSIS) Framework to optimize Gavi HSIS support for sustainable coverage and equity improvements. The HSIS Framework allows countries implementing HSS grants to access 100% of their HSS grant ceiling for all five years[16] (rather than 80% of the ceiling for years 2-5, as discussed above and shown in Figure 1). As applicable, PBF performance payments will be provided as *supplemental* funding on top of the HSS ceiling up to a value that cannot exceed 150% of the annual ceiling (calculated as one fifth of the five-year ceiling generated by the HSS Resource Allocation Formula). As a result of this change,

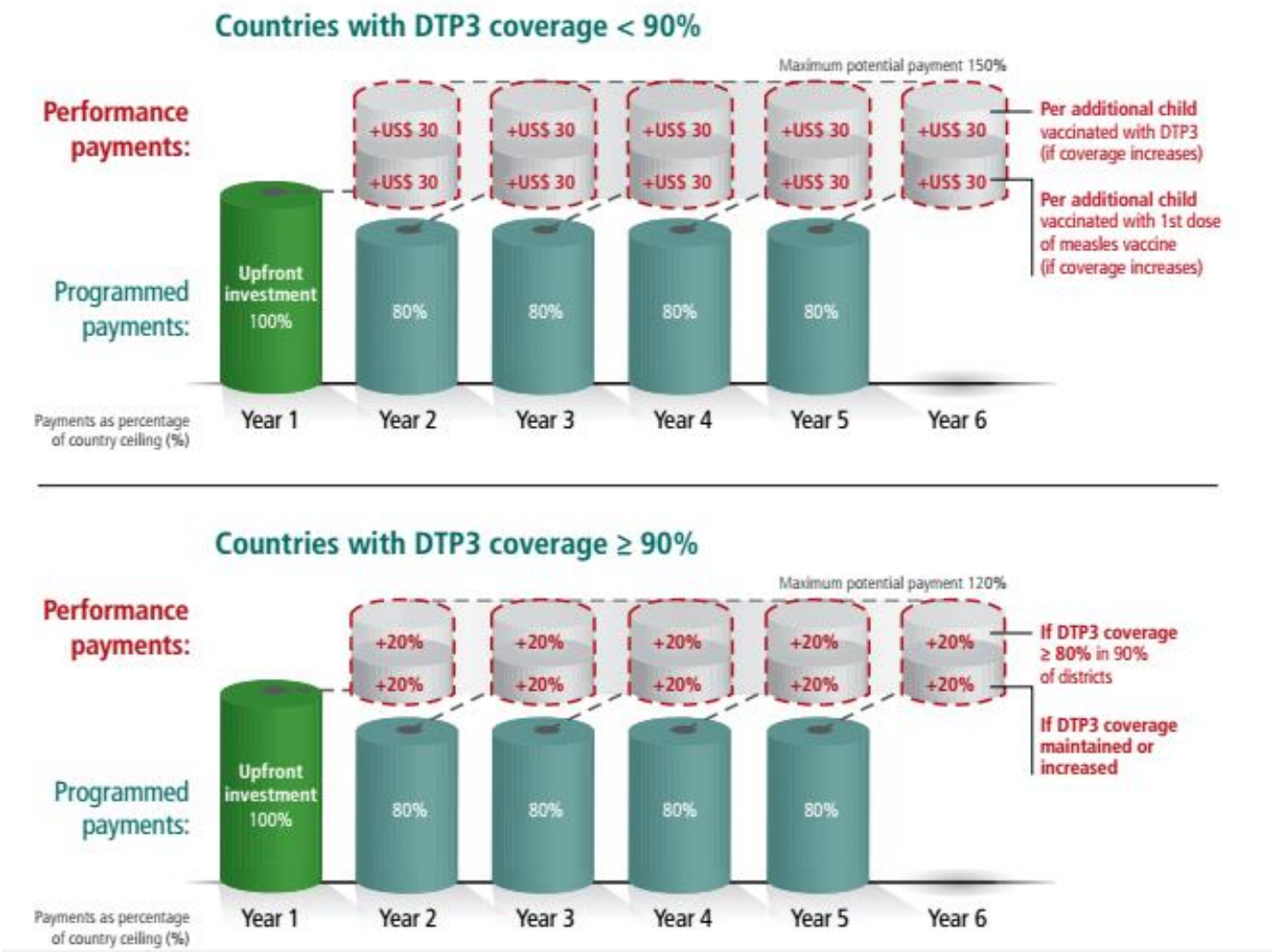
[13] Gavi's current strategy has four goals, each supporting Gavi's mission "to save children's lives and protect people's health by increasing equitable use of vaccines in lower-income countries." Goal 2 is the "systems goal" – it aims to "increase effectiveness and efficiency of immunisation delivery as an integrated part of strengthened health systems."
[14] Performance Based Funding Information Sheet, available at: www.gavi.org/library/gavi-documents/guidelines-and-forms/performance-based-funding-information-sheet/.
[15] If a country's total grant budget is under US$3 million, it is exempt from the 80% rule.
[16] https://www.gavi.org/library/gavi-documents/policies/gavi-health-system-and-immunisation-strengthening-support-framework/ (page 7, box on Performance payments).

countries with an HSS grant recommended for approval by the Independent Review Committee (IRC) in 2015 or 2016 to be implemented within the current Strategic Period will be allowed to program for the full amount of their originally communicated HSS ceiling. Thus, with the new approach, performance payments are no longer calculated as a part of the country's HSS ceiling, they are entirely supplemental to the HSS grant (Gavi will no longer "withhold" a portion of the country ceiling). This change affects the amounts that can be gained through receiving performance payments. Prior to this change, countries with DTP3 coverage equal to or higher than 90% at baseline (according to the WUENIC estimates) could obtain 20% of the annual ceiling for maintaining or increasing DTP3 coverage, and 20% for improving/maintaining a high geographic equity in DTP3 coverage. For the set of countries recommended for approval by the IRC in 2015 and 2016 that will be accessing 100% of their ceiling, the performance payment will only be worth 20% of their annual ceiling: 10% for maintaining/increasing DTP3 coverage and 10% for improving geographic equity. By doing this, high performing countries will still be getting a maximum potential payment of 120% of their ceilings.

► Performance payments have to be used in line with the guidance for HSS grants. As such, they need to be targeted "at interventions that improve coverage in under-immunized populations and areas".[17] These payments cannot be used to fulfil Gavi's co-financing requirements.

[17] Gavi, the Vaccine Alliance 2016: Gavi Alliance Health Systems and Immunisation Strengthening (HSIS) Framework, Gavi, the Vaccine Alliance 2018: Guidelines on Reporting and Renewal of Gavi support. As we discuss in this review, the guidance on the use of HSS funding changed over recent years.

**Figure 1. Gavi's PBF model for HSS cash support[18]**




[18] Figure from www.gavi.org/library/gavi-documents/guidelines-and-forms/performance-based-funding-information-sheet/. *Programmed payments* are based on progress in implementation and achievement of intermediate results. *Performance payments* are based on improvements in immunization outcome indicators.

**Table 1. Decision rules used by Gavi to determine whether a country is eligible for the performance payment**

| Countries with <90% DTP3 coverage at baseline* | Countries with ≥90% DTP3 coverage at baseline* |
|---|---|
| **DTP3 performance payment** | **DTP3 performance payment** |
| **If** DTP3 coverage is higher than the previous year or the previous highest coverage level since the initiation of the PBF grant, as per the country's administrative data,<br>DTP3 coverage as measured by the administrative data system is not more than five percentage points higher than the WHO/UNICEF estimate of DTP3 coverage for the country.<br>**Then** US$30 is awarded per additional child vaccinated with DTP3 relative to the previous year or the previous highest coverage level. | **If** DTP3 coverage is maintained at 90% or above, as per the country's administrative data,<br>DTP3 coverage as measured by the administrative data system is not more than five percentage points higher than the WHO/UNICEF estimate of DTP3 coverage for the country.<br>**Then** 20% (or 10%)* of the annual ceiling for the **HSS grant** is awarded. |
| **MCV1 performance payment** | **Equity performance payment** |
| **If** Routine measles first dose (MCV1) coverage is higher than the previous year or the previous high coverage level since the initiation of the PBF grant, as per the country's administrative data,<br>MCV1 coverage as measured by the administrative data system is not more than five percentage points higher than the WHO/UNICEF estimate of routine measles first dose coverage for the country requirement.<br>**Then** US$30 per additional child vaccinated with routine measles first dose relative to previous year or the previous highest coverage level since the initiation of the PBF grant. | **If** all districts in the country submitted immunization coverage reports (not to Gavi, but to the country's own reporting system),<br>**And** at least 90% of districts in the country have ≥80% DTP3 coverage, as per their reports,<br>**Then** 20% (or 10%)* of the annual ceiling for the **HSS grant** is awarded. |

*Depending on year of approval of HSS grant (see discussion on supplemental model on pages 7/8).

**Box 1: PBF measurement and verification**

Initially, PBF included only one mechanism for data verification (the "default option"): DTP3 and MCV1 country administrative data would be checked with the baseline estimate and the estimates for the previous year. [19] If coverage increases (or is maintained at 90% or above), the estimates are verified against the WUENIC estimates.[20] If estimates from country data are five or more percentage points higher than WUENIC estimates, the country would be ineligible for PBF.

To avoid penalizing countries with weak administrative systems, a new option (the "WUENIC" option) was introduced in 2015. Countries can choose to use the WUENIC data for their baselines and the annual PBF reviews. Countries will be eligible if coverage increases (or is maintained at 90% or above). At the same time, a third option was introduced. Countries can also conduct surveys to measure their coverage levels. [21]

Data verification options for calculating performance based funding payments are captured at the time of a country's HSS support application. Once an option is selected, it cannot be changed before a new HSS grant starts.

The only countries included in this review that requested an "alternative" data option for measuring their 2016 performance were Bangladesh (WUENIC option) and Pakistan (survey option). Going forward, there will certainly be additional countries that use these alternative options. A range of countries have new HSS grants (and will be assessed for PBF eligibility in 2019) and have selected the WUENIC option.[22]

The verification process is organized as follows: once a year, following the release of the WUENIC estimates (July 15th), Gavi's Application & Review (A&R) team conducts the eligibility assessment and the calculation of the performance payment. Gavi's High Level Review Panel (HLRP) officially approves the performance payment for high-impact countries. The Managing Directors approve the payment for non-high-impact countries. Countries must submit a budget to the Gavi Secretariat within three months of being communicated the performance payment amount for which they are eligible. This deadline will be communicated along with the performance payment eligibility, amount and timeline.[23] In practice, SCMs and POs are notified immediately by the A&R team, so that they can *informally* notify the country to allow for an initial discussion on the budget.

[19] Gavi, the Vaccine Alliance: Performance Based Funding (PBF) update for SF&P meeting. 29th August, 2017.
[20] http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tswucoveragedtp3.html
[21] Detailed information can be found in Guidelines for applications for HSS support in 2016; Section 4.2. - Data verification options for performance (WUENIC and Survey options).
[22] Afghanistan, CAR, Comoros, DRC; Eritrea, Liberia, Uganda, Zimbabwe (and Bangladesh in 2017).
[23] Gavi, the Vaccine Alliance 2018: Guidelines on Reporting and Renewal of Gavi support.

## 2. Methodology

In this section, we describe the overall evaluation methodology. We describe the evaluation framework (2.1) and the different methods that we used to collect and analyze data on Gavi's PBF approach (2.2).

### 2.1. Evaluation framework

We developed an **evaluation framework** to comprehensively assess the questions set out in the RFP. The framework broke down the overarching evaluation topics into more detailed questions to ensure that all relevant aspects were covered by the review. It also provided an overview of the methods for data collection and analysis to systematically assess the different questions and to measure the performance and results of the PBF approach (Annex 1).

The framework is based on the four evaluation dimensions set out in the RFP (design, implementation, results, and lessons learned) and the related evaluation questions:

►**Development and design:** One set of questions relates to the overall design of the PBF model: Is it fit for purpose? Does it have a sound theory of change? Does it suffer from any major design flaws? Were learnings from the ISS support taken into account? To assess the design, we established a set of criteria based on the literature review, a comparison of the Gavi PBF model with two related initiatives (the SMI and the Global Fund's PBF approach), and key informant interviews.

►**Implementation:** We differentiate between communication; budgeting, work planning, and grant-making; monitoring and evaluation (M&E); and harmonization with other results-based financing (RBF) schemes. We assessed PBF in the context of the overall Gavi HSS grants and analyzed the use of PBF payments. We also assessed whether the PBF approach has been implemented as anticipated or whether the implementation departed significantly from the originally planned process.

►**Results**: The results of the PBF are key for our recommendations on whether the PBF model should be continued and, if it should, how it should be redesigned. Assessing results of the PBF requires a "theory of change" (ToC), which lays out the underlying implementation logic and links the inputs to the intended outputs and outcomes (in this case, immunization coverage). As no ToC was developed for the current PBF model, we established a high-level version to assess the effectiveness of the model (Annex 2). The ToC was based on a document review (including two documents from the Gavi Task Team on HSS). We also discussed the ToC during our interviews with Gavi staff, global partners, and experts on PBF/RBF approaches.

Based on this ToC, we assessed the results of the PBF along three dimensions:

- **Ability to earn reward payments:** One important question is whether the PBF is in line with its stated intention "to improve immunisation coverage across countries by strengthening the underlying system," or whether it has failed to support countries evenly and, if so, why. To analyze this question, we conducted a financial assessment to understand (i) which countries earned performance payments, (ii) the likelihood that they received performance payments, and (iii) the reasons why countries failed to receive such payments.

- **Motivational aspect**: Important to this ToC is whether the PBF performance payment has a motivational effect on country behavior, i.e. whether it contributes to better immunization planning and implementation. Using results from key informant interviews, we thus analyzed the motivational effect of Gavi's PBF, i.e., whether it contributed to better planning and implementation, and whether it led to an overall prioritization of immunization coverage and equity within countries. Assessing whether the PBF model had a motivational effect is critical because an in-depth assessment of coverage trends is difficult and beyond the scope of this review.
- **Coverage and equity:** We also discussed in the interviews with country stakeholders whether they have any evidence to show that the PBF has contributed to improvements in coverage and equity. If countries did have evidence, we asked them to share these data to assess whether the PBF contributed to increases in coverage and equity. If countries did not have any hard data, we asked them if it is reasonable to assume that the PBF contributed to improvements in coverage and equity. Rather than providing definitive proof, the goal of this analysis is to provide a plausible "contribution story," which, based on available evidence and a line of reasoning, shows that the observed outcomes can potentially be attributed to Gavi's PBF policy (or shows that it cannot, or shows that it can but only to a certain extent). The country deep dives used all available data, including data shared by countries, on subnational coverage, to arrive at plausible assumptions of the contributions of Gavi's PBF on coverage and equity.

►**Lessons learned:** Based on a triangulation and synthesis of the different findings, we lay out the key lessons learned in order to develop recommendations on Gavi's PBF going forward. The triangulation uses a two-step approach. First, we conducted an assessment based on the literature review and the evaluation criteria used to analyze the design and implementation of the PBF model. Second, we analyzed to what extent results were confirmed or disconfirmed by the key informant interviews.

### 2.2 Methods

Our review of Gavi's PBF approach is based on (a) key informant interviews, (b) analysis of relevant documentation of the approach (e.g. from policy documents and memos), (c) analysis of available quantitative data (e.g. examination of whether the size of the performance payments and disbursements varied between countries with high versus low baseline DTP3 coverage), and (d) a review of the literature on the impact of health-related PBF schemes in low- and middle-income countries. We triangulated between these data sources to gain a comprehensive understanding of the performance and results of the PBF approach. As described below, our review also includes five short country case studies ("country deep dives").

**More specifically, we have:**

▪ **Conducted interviews with the Gavi secretariat** to understand the strengths/weaknesses of the PBF approach, the need to adjust the PBF policy, the main drivers behind critical policy decisions, and

alternative design options. These interviews have included members of the HSIS team, SCMs, POs, and others. In total, we interviewed 26 Secretariat members.

▪ **Interviewed former Gavi staff members**, including those who were closely involved with the original design of the PBF and its redesign, as well as **Gavi's global partners** (e.g. the Bill & Melinda Gates Foundation, the World Bank, and the WHO). Overall, we interviewed 17 alliance partners and former Gavi staff.

▪ **Interviewed country level stakeholders:** We have interviewed 31 country level stakeholders from 17 countries (Annex 3 includes an overview of all interviewed key informants). We approached all 31 countries that were assessed for eligibility between 2014 and 2017 except Bangladesh, Cambodia, Comoros, Djibouti, Liberia, and Papua New Guinea. The Gavi Secretariat indicated that the timing of the review is difficult for these countries as they are currently working on other critically important Gavi requests, or they just underwent significant staff changes and as such lack knowledge on PBF.

▪ **Reviewed relevant written materials**, including strategic and policy documents from the Gavi Board and PPC, internal memos, and Gavi evaluations.

▪ **Analyzed financial data** on eligibility, performance payments, and disbursements.

▪ **Carried out a rapid benchmarking exercise** to compare Gavi's PBF approach with that of the Global Fund and the SMI.

▪ **Reviewed the key literature on PBF and RBF** for health improvement in low- and middle-income countries, to establish criteria for assessing the design of Gavi's PBF.

▪ **Conducted five short country deep dives.**[24] In addition to the phone interviews with selected in-country stakeholders, we also carried out a deeper assessment of any relevant country data that we were able to collect during the evaluation timeframe (Annex 4). In proposing countries for consideration in our review, we used the following criteria:

- Among those countries that qualified for performance payments, inclusion of both countries with DTP3 coverage $\geq$ 90% at baseline and countries with DTP3 coverage < 90% at baseline;
- Inclusion of at least one country that consistently failed to qualify for a reward payment (so that we could compare successful versus unsuccessful countries);
- Inclusion of fragile and non-fragile states;
- Variation in geographic region;
- Variation in country population size;
- Inclusion of French-speaking, and English-speaking countries.

Based on these criteria, we selected the five countries shown in Table 2.

[24] We initially planned to also develop a deep dive of Nicaragua but it was not possible to conduct interview due to the political crisis in the country.

**Table 2. Characteristics of the six countries proposed for the deep dive study**

| Country | Allocated performance payments (USD, millions) | Year first qualified for performance payment | DTP3 coverage (WUENIC) at baseline[1] | Region | Population size, millions[2] | Fragile state[3] |
|---|---|---|---|---|---|---|
| Burundi | 4.3 | 2014 | High (96%) | Sub-Saharan Africa | 10.5 | Yes |
| Ethiopia | Never qualified for reward | N/A | Low (69%) | Sub-Saharan Africa | 102.4 | No[3] |
| Lao PDR | 2.3 | 2014 | Low (78%) | South Asia | 6.8 | No |
| Sudan | 4.7 | 2015 | High (93%) | Sub-Saharan Africa | 39.6 | Yes |
| Tanzania | 4.0 | 2015 | High (91%) | Sub-Saharan Africa | 55.6 | No |

[1] Countries with DTP3 coverage levels at or above 90% are considered high coverage countries while those with DTP3 coverage levels less than 90% are considered low coverage countries.
[2] Based on World Bank data, 2016 (https://data.worldbank.org/indicator/SP.POP.TOTL).
[3] Based on Gavi's 2018 fragility list. At the time of selecting the countries, Ethiopia was on Gavi's 2017 list of fragile states.

This range of approaches helped to ensure that a broad variety of perspectives were included in the review, as highlighted in Gavi's evaluation policy. Based on these analytic approaches, we then developed a set of initial policy recommendations/options related to the future of Gavi's PBF scheme. These initial recommendations were discussed with Gavi staff during a focus group meeting in Geneva on August 23, 2018. Based on this discussion, we revised the recommendations.

### 2.3. Limitations

It is not possible as part of this evaluation to rigorously assess the extent to which Gavi's PBF has led to increases in routine vaccination coverage and equity. Contributions to increases in coverage are difficult to assess given that multiple factors impact on coverage levels. A rigorous assessment of the effect of Gavi's PBF on immunization coverage would have required a very different evaluation design, i.e., assessments at country level (e.g. surveys), including control groups and randomization to control for alternative explanations (confounding variables). Definitive proof that Gavi's PBF contributed to improvements in coverage and equity can therefore not be provided as part of this review. While validated through SCMs and assessment of available quantitative coverage data, some of the information is self-reported by country managers.

## 3. Criteria to assess the PBF design

In this section, we establish criteria for assessing the design of Gavi's PBF based on three approaches:

- First, in section 3.1, we summarize the literature on previous PBF schemes. While the overall effectiveness of PBF for improving uptake of health services in low- and middle-income countries remains unclear, nevertheless there have been enough evaluations to begin defining general factors associated with increased chances of success. In addition, we note additional criteria associated with success when designing global health programs more generally, many of which were recurring themes emerging from the key informant interviews.
- Next, we summarize the evidence on two large scale PBF schemes that are particularly relevant to Gavi's PBF model, as they share similar features—the SMI and the Global Fund's PBF approach (section 3.2). Several key informants mentioned the importance of comparing Gavi's PBF with these two other schemes, so section 3.2 includes a comparative analysis.
- Finally, in section 3.3, we draw on the findings of sections 3.1 and 3.2 to establish a selected set of criteria against which we then evaluate the design of Gavi's PBF.

### 3.1 What do existing studies say about the effectiveness of performance-based programs in global health?

**Key findings: Since the effectiveness and cost-effectiveness of PBF for improving global health outcomes remain unclear, PBF should still be considered an "experimental" approach and its use should always be accompanied by evaluation and learning. Unintended consequences and high transaction costs are well documented. Weaker countries can find it harder to qualify for reward payments, potentially worsening between-country inequities.**

There have been multiple systematic and non-systematic reviews of the evidence on using PBF to improve health service delivery in low- and middle-income countries; we summarize six of these reviews.[25] They have all come to very similar conclusions: **the overall effectiveness of PBF remains unclear**. Some studies have shown positive impacts, some have shown no impacts, and some have shown harms. So PBF should still be considered as an "experimental" approach, with unclear effects, making it all the more important that its use should always be accompanied by evaluation and learning.

Published evidence syntheses have used different criteria for including or excluding studies—thus they are each examining a different set of studies (though there is some overlap). It was not possible for us to examine every published review of PBF within the timeframe of this project; instead, we summarize six of the most prominent evidence summaries (three are systematic reviews, three are narrative reviews). Table 3 summarizes the findings of the six reviews. Box 2 defines key terms used in this review.

[25] The most reliable and rigorous syntheses are systematic reviews—these have pre-defined criteria for inclusion/exclusion and are less prone to bias than non-systematic (narrative) reviews.

**Table 3. Summary of six reviews of PBF and RBF schemes that aimed to improve health service delivery in LICs and MICs**

| Authors (Year) | Countries included in review | Review type and study types included in review | Intervention assessed in review | Key outcomes |
|---|---|---|---|---|
| Witter et al (2012) | Burundi, China, DRC, Philippines, Rwanda, Tanzania, Vietnam, Zambia | Systematic review. Studies were only included if they were RCTs, CBAs, or ITS studies | Payments for reaching target coverage, CCTs | ▪ Mixed picture of success: overall, studies showed incentive was associated with increased coverage of *some* but not all specified indicators |
| Das et al (2016) | Burundi, DRC, Egypt, Philippines, Rwanda | Systematic review. Included cluster RCTs, CBAs, and a case control study with a post-intervention comparison | Performance payments to facilities if achieve certain MCH performance targets | ▪ PBF was associated with improved "process quality" (adherence to protocols etc.)<br>▪ Mixed evidence on impacts on health outcomes: some studies showed improvements, others showed none |
| Oxman and Fretheim (2009) | Very wide range of LICs, MICs, and HICs (including LICs/MICs in sub-Saharan Africa, Latin America, Asia) | Review of 12 systematic reviews | CCTs, performance payments targeting health workers, or organizations | ▪ Incentives targeting individuals (patients, providers) are effective in short run for simple, distinct, well-defined behavioral goals.<br>▪ Little evidence that incentives can lead to sustained, system-wide improvements |
| Pearson et al (2010) | Very wide range of mostly LICs and MICs (though also examined P4P schemes in HICs) | Narrative (non-systematic) review; all study types | Results-based aid in LICs/MICs (e.g. Global Fund, Gavi ISS), RBF in LICs/MICs (e.g., Health Results Innovation Trust Fund), RBF in HICs (e.g. Medicare P4P schemes). | ▪ PBF schemes in LICs and MICs can "deliver results" but evidence base is generally weak and it is difficult to attribute positive results specifically to results-based focus rather than complementary reforms<br>▪ Unintended consequences are widespread and costs are large<br>▪ Performance rewards can worsen equity *between* countries (weaker countries are *less* likely to receive reward) but no evidence that such rewards worsen *within-country equity* |
| Paul et al (2018) | Wide range of LICs and MICs | Narrative (non-systematic) review; all study types | PBF programs (not further specified) | ▪ Overall, evidence remains unclear on effectiveness of PBF, its cost-effectiveness, and its impacts on equity<br>▪ Costs are high, related to managing the program and generating verification data |
| Perakis & Savedoff (2015) | Gavi-eligible countries, countries included in Salud Mesoamérica Initiative | Narrative (non-systematic) review; all study types | Results-based aid | ▪ RBA is a relatively underutilized and untested way of providing funding to governments<br>▪ When RBA programs work, it is unlikely to be due to the financial incentive, but to the increased results focus by politicians and bureaucrats |

Abbreviations: RCT: randomized controlled trial; CBA: controlled before-and-after; ITS: interrupted time series, CCT: conditional cash transfer, CBA: controlled before-and-after; MCH: maternal and child health

**Box 2: Defining terms used in the review**

There are many different definitions of PBF and the related terms pay-for-performance (P4P) and RBF. The most relevant definition of PBF for our review is *a mechanism by which incentive payments are earned on the basis of achieving specific pre-determined performance criteria that are measurable and verifiable*.[26] Most studies of PBF have examined schemes in which the financial incentives are applied at the level of the health facility or health provider, but incentives can also be applied at the national level, as is the case with Gavi's PBF. Such performance programs that apply financial incentives at the national government level are sometimes called results-based aid (RBA). A 2015 evidence review by the Center for Global Development, called "Does Results-Based Aid Change Anything? argued that there are only two examples of RBA used *specifically to improve health outcomes*: Gavi's ISS multi-country program and Salud Mesoamérica.[27] Other examples of RBA addressed deforestation and education.

The term PBF (which is synonymous with P4P) usually refers to **supply-side financial incentives**. Examples of **demand-side financial incentives** include conditional cash transfers (CCTs) and vouchers given to mothers to attend antenatal clinics. The umbrella term RBF covers all these various forms of incentives. We mostly focus on supply-side incentives, though we have mentioned demand-side incentives (e.g. CCTs) when they are relevant to our review.

One of the challenges in examining the published literature on PBF is that it is not a homogeneous intervention. As Logan Brenzel and colleagues noted in their February 12, 2014 paper for the TAG-HSS: "Performance-based financing is not a uniform intervention, but rather a range of approaches….the effects of these mechanisms depend on the intervention design, amount of additional funding, level of technical support, and the organizational context in which the mechanism is implemented."[28]

Below we have distilled the key messages of each of the six reviews. Additional details of these reviews are in Annex 5.

▪ **The 2012 Cochrane review by Witter et al: the effects of PBF on health care and health outcomes in LICs and MICs:** This systematic review,[29] which we believe is the most rigorous conducted to date (it only included controlled studies or interrupted time series studies), concluded that "the current evidence base is too weak to draw general conclusions; more robust and also comprehensive studies are needed."[30]

[26] Renmans D, et al. Opening the 'black box' of performance-based financing in low- and lower middle-income countries: a review of the literature. *Health Policy and Planning* 2016;31: 1297-1309.
[27] https://www.cgdev.org/publication/does-results-based-aid-change-anything-pecuniary-interests-attention-accountability-and
[28] Brenzel L, et al. Investing in Immunization Outcomes with GAVI HSS. Background paper for TAG-HSS, Feb 12 2014.
[29] Systematic reviews are considered to be the gold standard for evidence on health interventions and are at the very top of the "hierarchy of evidence."
[30] Witter S, Fretheim A, Kessy FL, Lindahl AK. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev.* 2012 Feb 15;(2):CD007899.

▪ **The 2016 systematic review by Das et al: the effects of PBF on the quality of maternal and child health care in LICs and MICs:** The review concluded that there is only "weak evidence for P4P's positive effect on maternal and neonatal health outcomes and out-of-pocket expenses."[31] The review found evidence that P4P can improve "process quality" (e.g. adherence to guidelines).

▪ **Oxman and Fretheim's 2009 "review of reviews": a review of 12 systematic reviews of the health sector effects of any type of RBF in LICs and MICs:** The authors found that financial incentives targeting patients and health providers can be effective in the short term for simple, distinct, targeted behavioral goals, but it is unclear whether changes can be sustained in the long term.[32] CCTs can improve uptake of preventive health services. Other than these findings there is "very limited evidence of the effects of results-based financing in low- or middle-income countries." There is evidence that it can cause harms (e.g. worsening of health disparities).

▪ **DFID's 2010 review of results-based aid and results-based financing schemes:** Mark Pearson and colleagues in DFID's Human Development Resource Centre reviewed RBA and RBF schemes in LICs and MICs and concluded that they can "deliver results" (i.e. some schemes have been associated with improved outcomes), but the evidence base is generally weak.[33] Transaction costs can be large, especially the costs of supervising the PBF scheme and of monitoring and measuring progress. Unintended effects are widespread, such as a worsening of the quality of services in order to achieve a quantitative numerical PBF output target. PBF can worsen equity *between* countries—for example, in the PBF scheme used by both the Millennium Challenge Account and Gavi's ISS, weaker countries were less likely to receive the rewards, which can then worsen between-country equity. Although there have been concerns that the poor *within* countries will be disadvantaged and left behind by PBF schemes, Pearson and colleagues' review concluded that "this is not inevitable and fears that it would occur have often proven to be largely unfounded."

▪ **Paul and colleagues' 2018 review: evidence on problems with PBF schemes:** Paul and colleagues reviewed documented problems with the use of PBF schemes for improving health outcomes in LICs and MICs, focusing in particular on PBF programs supported by the Health Results Innovation Trust Fund (HRITF). The review pointed to the lack of high quality, credible, reproducible evidence on the effectiveness of PBF, its cost-effectiveness, or its impacts on equity.[34] Paul and colleagues found evidence of high transaction costs in several PBF programs, related to managing the program and generating verification data. For example, in Tanzania, PBF economic cost (e.g. the cost of time spent on verification) was double the financial cost;[35] in Benin's PBF scheme, supported by the World Bank, for every US$1 paid to providers, US$0.50 is used for verification;[36] and in Burkina Faso's PBF scheme

[31] Das A, Gopalan SS, Chandramohan D. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. *BMC Public Health* 2016;16: 321

[32] Oxman AD, Fretheim A. Can paying for results help to achieve the Millennium Development Goals? Overview of the effectiveness of results-based financing. *J Evid Based Med.* 2009 May;2(2):70-83

[33] Pearson M, Johnson M, Ellison R. Review of Major Results Based Aid (RBA) and Results Based Financing Schemes. DFID Human Development Resource Centre, 2010. At https://www.oecd.org/dac/peer-reviews/Review-of-Major-RBA-and-RBF-Schemes.pdf

[34] Paul E, Albert L, Bisala BN, et al. Performance-based financing in low-income and middle-income countries: isn't it time for a rethink? *BMJ Glob Health* 2018; 3(1): e000664.

over 2014-2015, 30% of the total costs of PBF were for operations.[37] Paul and colleagues also found evidence of PBF schemes being scaled up even when pilots were negative. The World Bank's Independent Evaluation Group (IEG) found that multiple PBF pilots had failed and yet "decisions were made to scale up regardless of weak, inconclusive, or incomplete pilot results."[38] The IEG gave the examples of Benin, Tanzania, and Argentina where the Bank pushed for nationwide PBF despite lack of evidence from pilots. Lastly, the authors argue that use of PBF in LICs and MICs has been donor-driven, not home grown, and there is little in the way of domestic ownership.

▪ **The Center for Global Development (CGD) 2015 review of RBA schemes:** Rita Perakis and William Savedoff at CGD assessed six aid programs that paid governments for outcomes, of which two had a health focus (the other four focused on education and deforestation). These programs all showed some success, but this was unlikely to be due to the financial incentive. Instead, RBA programs "seem designed to draw attention to results, making them more salient to politicians and managers." This review found that concerns about RBA, such as corruption, unintended consequences, and a focus on the short term over the long term have largely not materialized.

**Key findings: A number of design features have been associated with greater chances of success for a PBF program. Examples include setting and communicating clear and simple objectives, targets and indicators; assessing institutional and organizational capacity up front; strong measurement and verification systems; the use of larger reward payments; and sufficient autonomy in how the performance payment is used.**

▪ **Setting and communicating clear and simple objectives, targets, and indicators.** An important determinant of PBF program performance is whether the objectives, targets, and indicators are simple and clear enough and whether the potential recipients of the performance payment (e.g. health facilities) understand them. Precise and specific targets help to clarify tasks and responsibilities.[39] The HRITF's PBF conceptual framework, for example, which lays out determinants of success and interlinkages between them, notes the importance of the health facility's "understanding" (defined as "the knowledge of criteria by which incentives are awarded, the amount of money at stake, and the additional design features"[40]).

▪ **Robust measurement and verification systems.** Pearson and colleagues note the importance of paying attention to performance data and data verification up front, arguing that explicit attention must be paid in the scheme's design to address this issue. Robust, reliable measurement and verification processes are the cornerstone of PBF. Verifying whether performance targets have been met is fundamental to ensuring transparency and the long-term viability of PBF programs. Nevertheless, there is a trade-off between validity and costs—PBF programs need to find the right balance that provides robust enough verification at reasonable cost.

[39] Renmans D, et al. Opening the 'black box' of performance-based financing in low- and lower middle-income countries: a review of the literature. *Health Policy and Planning* 2016;31:1297-1309.
[40] https://www.rbfhealth.org/resource/performance-based-financing-conceptual-framework

▪ **The size of the reward payment.** There is some evidence from empirical studies that larger rewards are associated with increased odds of successful outcomes. For example, in a randomized controlled trial in Rwanda, Paulin and colleagues randomly assigned facilities at the district level either to begin pay-for-performance (P4P) funding or to continue with traditional input-based funding. [41] P4P was associated with higher use and quality of several maternal and child health care services, but it had no effect on use of antenatal care or on the timely completion of child immunization schedules. The study found that the effects were larger for those services "for which facilities receive larger financial incentives and those over which the provider has greater control." Early evidence from impact evaluations of HRITF programs also suggests that the performance payment needs to be large enough to incentivize changed behavior.[42]

▪ **The speed of the reward payment.** There is evidence that the reward is more likely to have a motivating effect if it is received soon after the achievement of the target that "triggers" the payment. Long delays in receiving the payment, and poor predictability of the reward, can hinder success.

▪ **Autonomy and a degree of flexibility in how the reward payment is used**. There is evidence showing that reward payments are more likely to be effective when providers have greater autonomy over how it is used (as previously discussed, such autonomy is a feature of the SMI). Initial results from several PBF programs supported by the HRITF also found that financial autonomy was associated with a number of benefits. For example, in Argentina's *Plan Nacer* program, which uses PBF to improve the use and quality of maternal and child health services, evaluation found that "the financial autonomy provided to facilities by *Plan Nacer* allowed a better allocation of scare resources, which in turn had a positive impact on health outcomes of beneficiaries."[43]

▪ **Adopting a differentiated, country-specific approach**. When it comes to the design of a PBF program, studies suggest that there is no "one size fits all" approach that will be suited to every different country context. As Renmans and colleagues say, "as for any other programme in the health sector the context is an important factor in the effectiveness and appropriateness of a PBF scheme."[44] PBF schemes are not rolled out in isolation or in a vacuum—they depend strongly on country capacity and the strength of the health system (e.g. the health management and information system, financial management capacity, etc.). Thus PBF approaches should be tailored to different contexts, based on an assessment of a country's readiness. Without such tailoring, PBF schemes risk worsening equity between countries. Using PBF models in fragile, conflict affected states (FCAS) can be particularly challenging. Several studies, including the 5-year evaluation of Gavi's ISS and the IHME

[41] Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *Lancet* 2011; 377:1421-8.
[42] https://www.rbfhealth.org/resource/completed-impact-evaluations-and-emerging-lessons-health-results-innovation-trust-fund
[43] https://www.rbfhealth.org/resource/completed-impact-evaluations-and-emerging-lessons-health-results-innovation-trust-fund
[44] Renmans D, et al. Opening the 'black box' of performance-based financing in low- and lower middle-income countries: a review of the literature. *Health Policy and Planning* 2016;31: 1297-1309.

evaluation of ISS,[45, 46] have shown that country fragility can be a major barrier to the success of PBF due to problems such as weak health information systems, governance, and leadership. Global Fund grants in FCAS settings have performed worse than in stable countries.[47] Multiple evaluations have also concluded that "a one size fits all performance-based scoring system (as used by the GAVI Alliance and the Global Fund) might not allow for the limited availability and quality in data quality inherent in many conflict-affected countries."[48]

One way to improve equity *between* countries when it comes to PBF is to provide weaker countries with (i) up front capacity building so that they can actually take advantage of PBF schemes, and (ii) simpler application and approval processes. Other approaches include providing larger performance payments for poorer countries and using "locally identified and relevant targets."[49]

▪ **Assessment of institutional and organizational capacity up front.** If countries do not have this capacity, up-front investments to tackle this bottleneck will be needed before a PBF scheme is instituted. Similarly, ongoing support (the so-called "ancillary components" of PBF schemes) is a key success factor in PBF schemes.[50] Coupling financial incentives with tailored technical assistance, supervision, and training can boost motivation and increase the likelihood of program success. The World Bank's PBF Toolkit notes that "continuous support during the early stages of introducing PBF—when people are still grappling with understanding the new system—is vital."[51]

▪ **Using PBF approaches within a learning environment.** As mentioned previously, there are many unanswered questions about PBF, including questions related to its effectiveness, cost-effectiveness, and the influence of contextual factors. PBF models should therefore combine implementation with learning (what Peters and colleagues call "learning and doing"[52]) and continually use the lessons to refine and improve implementation.

## 3.2 The Salud Mesoamérica Initiative (SMI) and the Global Fund's PBF approach

[45] Chee, G, Hsi, N, Carlson, K, Chankova, S, and Taylor, P. ((accessed Dec 15, 2008).)Evaluation of the first five years of GAVI Immunization Services Support funding. Abt Associates, Bethesda; 2007.
[46] Lu, C, Michaud, CM, Gakidou, E, Khan, K, and Murray, CJL. Effect of the Global Alliance for Vaccines and Immunisation on diphtheria, tetanus, and pertussis vaccine coverage: an independent assessment. Lancet. 2006; 368: 1088–1095
[47] Bornemisza O, Bridge J, Olszak-Olszewski M, Sakvarelidze G, Lazarus J. Health aid governance in fragile states: the Global Fund experience. *Global Health Governance*. 2010;4(1).
[48] Patel P, et al. Exploring the influence of the Global Fund and the GAVI Alliance on health systems in conflict-affected countries. *Confl Health* 2015; 9: 7.
[49] Pearson M, Johnson M, Ellison R. Review of Major Results Based Aid (RBA) and Results Based Financing Schemes. DFID Human Development Resource Centre, 2010. At *https://www.oecd.org/dac/peer-reviews/Review-of-Major-RBA-and-RBF-Schemes.pdf*
[50] Janssen W, Ngirabega Jde D, Matungwa M, Van Bastelaere S. 2015. Improving quality through performance-based financing in district hospitals in Rwanda between 2006 and 2010: a 5-year experience. Tropical Doctor 45: 27–35
[51] Fritsche, György Bèla, Robert Soeters, and Bruno Meessen. 2014. Performance-Based Financing Toolkit. Washington, DC: World Bank
[52] Peters DH, El-Saharty S, Janovsky K (2009) From evidence to learning and action. In: Peters D, El-Saharty S, Siadat B, Janovsky K, Vujicic M, editors. Improving health service delivery in developing countries. Washington, DC: World Bank. pp. 277–296.

**Key findings: A comparison of Gavi's PBF with the SMI found that the SMI has a more comprehensive set of supportive activities accompanying the reward payment (including technical assistance) and it gives countries greater autonomy in how the payment is used. The SMI's external verification mechanism is much more costly than Gavi's approach (the SMI's mechanism uses 14% of the donor funding). Both the SMI and Gavi's PBF have very short time frames for achieving the targets, which many countries find challenging. There are only a few examples of cash-on-delivery programs at the Global Fund. One notable feature of the Global Fund's PBF is that the Global Fund allows countries more time to achieve the PBF coverage targets than in other similar schemes.**

**The Salud Mesoamérica Initiative (SMI)**

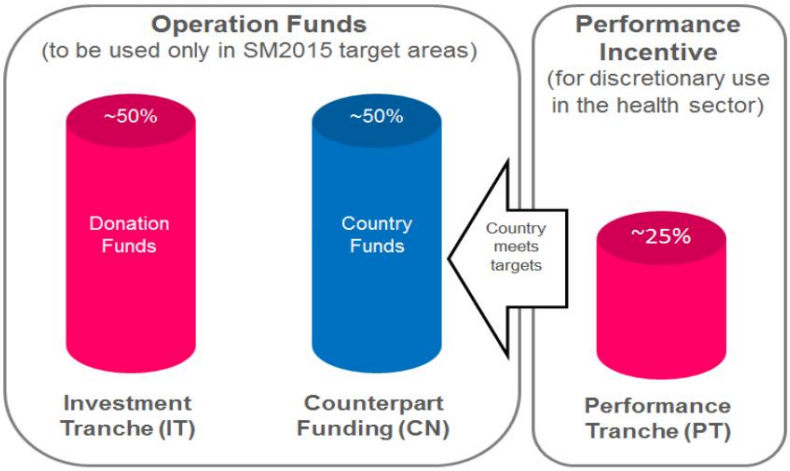
***Background***

The SMI is a PBF scheme administered by the Inter-American Development Bank (IDB) that aims to improve maternal and child health (MCH) among the poorest quintile of the population in eight countries: El Salvador, Guatemala, Honduras, Nicaragua, Belize, Costa Rica, Panama, and the State of Chiapas, Mexico. It is funded by three donors—the Carlos Slim Foundation, the Gates Foundation, and Spain's Cooperative Agency for International Development—and the ministries of health of the eight participating countries. The program itself involves the scaling up of evidence-based MCH interventions (e.g., family planning, antenatal and postnatal care, facility-based deliveries, childhood vaccination) and health systems improvements (e.g., improving supply chains).

***Design of the SMI's performance-based scheme***

For each project there is a set of performance (payment) indicators and targets (e.g. contraceptive coverage, complete vaccination coverage for age). Half of the funding for operations comes from donors (the "investment tranche") and the other half from counterpart funding from countries themselves. When the operation finishes, country performance is independently measured and if 80% of facilities meet all of the predefined indicators, the country receives a reward payment, valued at half of the counterpart funding (this is the "performance tranche") (Figure 2). Total funding for the SMI is US$169 million (US$114m from donors, US$55m from domestic funds).

**Figure 2. Design of the performance payment in the SMI**. Figure from the IDB.




***Initial evaluation of the SMI***

The first evaluation of SMI was published in April, 2018, by Mokdad and colleagues, focusing on the facility supply of essential medicines and services (*not* coverage, quality of interventions, or health outcomes).[53] Baseline data were collected from 365 intervention health facilities and follow-up was at 18-24 months post-intervention. Five countries reached their predetermined performance targets and received the performance tranche (El Salvador, Honduras, Nicaragua, and Panama). The other three countries showed progress but failed to reach the targets and thus did not qualify for the reward payment. Chiapas and Guatemala instituted a performance improvement plan (using their own funds) and achieved the previously missed targets 9-12 months later.

**Comparison with Gavi PBF**

We drew on Mokdad and colleagues' evaluation, together with a number of IDB reports,[54] to compare the SMI against Gavi PBF on four domains (Table 4):

**Table 4. Comparison of the SMI with Gavi's PBF**

[53] Mokdad AH, Palmisano EB, Zúñiga-Brenes P, Ríos-Zertuche D, Johanns CK, Schaefer A, et al. (2018) Supply-side interventions to improve health: Findings from the Salud Mesoamérica Initiative. *PLoS ONE* 13(4): e0195292

[54] See, for example, Iriarte E, et al. The Initial Prize in the Salud Mesoamerica Initiative Results-Based Aid Initiative. IDB Technical Note No. IDB-TN-1314. October 2017.

| Domain | SMI | Gavi PBF | Implications |
|---|---|---|---|
| *Activities that complement reward payment* | ▪ Intensive set of complementary activities, including country-specific technical assistance to overcome bottlenecks<br>▪ SMI works with countries that fail to achieve reward on a "performance improvement plan" | ▪ PBF approach is less comprehensive, with narrower focus on the reward payment itself<br>▪ No specific approach to helping poor-performing countries increase future chances of success | ▪ Factors beyond the reward itself help to improve outcomes |
| *Verification* | ▪ External verification by IHME (baselines and periodic measurements)<br>▪ SMI chose to use external measurement due to lack of credible health data on poorest quintile and to build donor confidence<br>▪ Costs are significant: 14% of the US$115m is for measurement | ▪ Gavi's PBF scheme now gives countries three options: country administrative data verified against WUENIC data ("default option"), WUENIC data from the start ("WUENIC option"), or country surveys | ▪ External verification for Gavi PBF would be expensive. There is a trade-off between high validity of verification method and costs.<br>▪ Donor tolerance of risks influences choice of verification scheme |
| *Use of reward payment* | ▪ Countries have the autonomy to use the performance tranche for any health sector use | ▪ Payment must be used to fund "interventions that improve coverage in under-immunized populations and areas" | ▪ A sufficient degree of autonomy of the incentive recipient (provider, health facility, or country government) is a key success factor |
| *Time frame* | ▪ 18-24 months to reach performance targets | ▪ Performance payment is awarded annually, i.e. time frame of 12 months to achieve target | ▪ Initial evaluation of SMI scheme found that countries that failed to qualify for the payment found short time frame to reach targets very challenging. Gavi's time frame for PBF is even shorter |

**The Global Fund's PBF approach**

***Background***

The Global Fund's PBF approach shares key similarities with Gavi's general cash-based support. Under the principle of the Global Fund's PBF, continued grant funding is dependent on proven results. All grants have M&E frameworks, which are externally reviewed when they are created, and also include input, outcome, and impact indicators in addition to information on activities, expenditures, and grant management. As part of their annual progress update and disbursement request, recipients report on progress, which is verified by local fund agents.[55]

***Cash-on-delivery grants***

In addition to this overall approach, the Global Fund has specific results-based financing ("cash-on delivery") grants that are based on tying financing levels to the achievement of a smaller number of high-level impact or outcome indicators. In 2014, a cash on delivery grant was introduced in Rwanda. Disbursements are strictly based on the performance of six coverage indicators, which are included in the country's national strategic plan. Funds were pooled with the support from other development partners.

Since 2014, the Global Fund has supported the Elimination of Malaria in Meso America and Hispaniola Initiative. Using an innovative cash-on-delivery approach, the grant aims to reach zero malaria cases in nine countries by 2020 and to seek certification of malaria elimination by 2025. The cash on delivery program rewards progress by linking achievements to one single indicator, the number of local malaria cases, with 70% % of funds being performance-based. Countries were only rewarded post facto when impact was achieved and confirmed. An evaluation of this grant determined that the countries in Central America and Hispaniola have made significant progress in reducing malaria, with cases declining by 90% in the past two decades. However, the evaluation determined that the region is not currently on the trajectory needed to achieve the goal of elimination by 2020.[56] Overall, there are only a few examples of cash-on-delivery programs at the Global Fund. One notable feature of the Global Fund's PBF is that Global Fund allows for more time to achieve the PBF coverage targets than in other similar schemes.

## 3.3. Criteria for assessing design of Gavi's PBF

We used the findings of sections 3.1 and 3.2 to establish a set of seven criteria against which we evaluate the design of Gavi's PBF in Section 4.2. The final set of criteria are also influenced by our own observations and the discussions with key stakeholders involved in the development and implementation of Gavi's PBF model. The first six criteria are linked to the assumptions that underpin the PBF model, i.e. its implicit, underlying theory of change. The seventh criterion is broader, and more focused on the context in which the PBF model is embedded ("being embedded in a learning environment").

[55] The Global Fund to Fight AIDS, TB and Malaria: Revised Progress Update and Disbursement Request. March 2016.
[56] Building on this grant, the Inter-American Development Bank is creating a Malaria Elimination Blending facility, a multi-donor trust fund, to continue the work.

**Table 5: Criteria to assess the design of the PBF model**

| Criteria | Questions |
|---|---|
| Achievability of performance targets | Are the assumptions on reaching the performance indicators realistic? Are the goals achievable in the given timeframe, or are they overly ambitious? |
| Robust measurement and verification approach | Is the measurement approach sound? Can country performance accurately be measured and verified? What are the cost implications of the measurement? |
| Size of the performance payment | Is the size of the performance payment significant enough to incentivize countries to give more attention to immunization coverage and equity? |
| Speed of performance payment | Is the timing of the performance payment closely linked to the performance itself? Is the payment received quickly after? |
| Flexible use of performance reward | Is there a degree of flexibility or autonomy in how countries can use the performance payment? |
| Equity across countries | Does the PBF model consider the specific challenges faced by countries with very low capacity, and those affected by conflict and humanitarian emergencies? Does it provide equal access across countries? |
| Being embedded in a learning environment | Is the PBF model being used within a learning environment? |

# 4. Development and design of Gavi's performance-based funding model

## 4.1. Development of Gavi's PBF model

**This section addresses the following questions:**

- To what extent did the PBF design consider lessons learned from ISS?
- Was the PBF design sufficiently informed by country stakeholder consultations?
- How and to what extent did the PBF model meet the requests from Gavi's Board?
- How did the PBF design change over time?
- From which initial shortcomings did the PBF approach suffer and how were these issues addressed?

***Development of the PBF model in 2010/11***

Gavi's original PBF program, a component of the **Immunization Support Strengthening** (ISS) support, was introduced under Gavi's Phase 1 Strategy (2000–2006). Country applications were approved for five years of support and ISS funding was paid in instalments over three years, based on each country's self-projected number of children to be immunized with three doses of DTP3 in the first year after application. The reward funding was calculated at US$20 per additional child receiving DTP3 above the number of children targeted the first year after application. The system for reporting the number of children immunized with DTP3 was validated through a one-time Data Quality Audit (DQA) conducted by Gavi-retained external auditors.

The ISS support was evaluated in 2007, and the evaluation concluded that "GAVI ISS funding has been successful in achieving its stated goal of improving access to immunizations."[57] The evaluation recommended that GAVI should continue to provide ISS funding, but also made a series of recommendations on how to improve the model further. One recommendation was to increase the size of the payment.

In 2008, the Institute for Health Metrics and Evaluation (IHME) published a study on ISS, which found substantial over-reporting of immunization data. It presented evidence that "DQA results are an inadequate measure for assessing the validity of administrative data relative to surveys."[58] Following the publication of this study, GAVI conducted a review of ISS to assess the quality of data and resulting coverage estimates. The review recommended that Gavi continue to implement ISS.[59]

However, key Gavi donors raised concerns about the lack of a robust verification mechanism to avoid overreporting of coverage. In December 2009, the PBF-TT was convened in response to two requests from the Gavi Alliance Board:

[57] Grace Chee, Natasha Hsi, Kenneth Carlson, Slavea Chankova, Patricia Taylor. September 2007. Evaluation of the First Five Years' of GAVI Immunization Services Support Funding. Bethesda, MD: Abt Associates Inc.
[58] Lim SS, et al. Tracking progress towards universal childhood immunisation and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage. *Lancet.* 2008;372:2031-46.
[59] https://www.gavi.org/library/news/statements/2009/update-on-immunisation-services-support-(iss)/; https://www.gavi.org/library/news/statements/2009/immunisation-services-support-(iss)-update/

- To identify means of supporting countries with less than 70% DTP3 coverage to increase routine immunization coverage; and
- To identify opportunities for incorporating PBF into Gavi support to countries.

In December 2010, the Board also requested the Secretariat to establish mechanisms to ensure that GAVI funding through cash-based programs is designed to have a reasonable and demonstrable impact on immunization programs in the context of integrated service delivery, and that immunization *coverage* is a credible outcome indicator for these activities.[60]

The PBF-TT – together with Gavi's Policy & Performance Team – developed a new PBF model, which was accepted by the PPC and presented to the Board in November 2011.[61] This model included a split of funds into two different types of payments. For the first year, countries would receive all funds from GAVI on a fixed basis. After the first year, the fixed and performance payments would be provided annually if the necessary conditions have been met. This PBF model involved three country groups, for which the performance-based share of funding differed (Table 6). Coverage data provided for the performance assessment needed to be verified based on WUENIC data.

**Table 6: PBF model approved by the Board in November 2011**

| Country category | Fixed payment | Performance payment |
|---|---|---|
| ≥90% DTP3 coverage at baseline | 20% of total indicative envelope received each year as a fixed payment | 40% of indicative envelope for maintaining or increasing DTP3 coverage and the remaining 40% for ensuring that 100% of districts have ≥80% DTP3 coverage |
| 70-89% DTP3 coverage at baseline | 40% of total indicative envelope received each year as a fixed payment | If DTP3 coverage increases, US$20 per additional child immunized with DTP3; plus US$20 per additional child immunized with first dose of measles containing vaccine, if measles coverage increases |
| <70% DTP3 coverage at baseline | 60% of total indicative envelope received each year as a fixed payment | If DTP3 coverage increases, US$20 per additional child immunized with DTP3; plus US$20 per additional child immunized with first dose of measles containing vaccine, if measles coverage increases |

[60] GAVI Alliance Board Meeting, 30 November – 1 December 2010, Kigali, Rwanda, Final Minutes.
[61] Report to the GAVI Alliance Board. Board-2011-Mtg-3-Doc 13. 16-17 November 2011.

The Gavi Board approved the PBF model in November 2011, following the recommendation by the PPC.[62] It also requested to roll out a performance-based component for all HSS grants approved in 2012 and onwards.[63] The Board also noted that there is continued need for more reliable data and stronger verification:

> "The Board noted that data quality is a matter of concern – not only inaccurate data but also the possibility of manipulation of data. The Board discussed the associated reputational risk for GAVI and suggested developing criteria for assessing data quality. It was recognised that more work is needed to assess and improve country data systems and to advance innovation in coverage estimation, such as through the use of biomarkers, and that such work is included in the business plan for 2012."[64]

The Board also reiterated the need for more work on fragile states and underperforming countries, and country tailored approaches. As highlighted above, the 2007 ISS evaluation had already suggested investigating alternative approaches for conflict-affected countries.

The approved model was responsive to the requests by the Board in that it focused on immunization outcomes. Key board members, such as the Bill & Melinda Gates Foundation, requested the inclusion of measles vaccine coverage into the model, which was included in the new model. The PBF model, to a certain extent, also considered the specific needs of countries with baseline DTP3 coverage levels below 70%: the lowest-coverage group received the highest proportion of payments on a fixed rather than performance basis, considering the more challenging country context. The PBF model also included an equity indicator for countries with baseline coverage of ≥90%, which became increasingly important to Gavi, as national coverage rates often mask inequalities in immunization coverage within countries.

At the same time, the PBF model kept the incentive payment at the same size as the former ISS support (US$20 per additional child immunized). As such it did not follow the recommendation of 2007 ISS review, which suggested increasing the amount paid per additional child vaccinated ("as US$20 may not provide sufficient incentive in countries with higher immunization coverage rates.").[65]

At the time of the Board decision, it was expected that all of Gavi' cash support would be channeled through the Health Systems Funding Platform.[66] The Platform was designed to improve the way Gavi, the Global fund, the World Bank, and other external funders support countries to strengthen their health systems, in alignment with jointly assessed national strategies and the principles set out in the Paris Declaration on Aid Effectiveness. However, the Platform was never roll out as planned.[67]

[62] The Board also decided not to proceed with IRIS [incentive approach for routine immunisation strengthening] as a standalone window of support. IRIS was piloted based on a Board decision from December 2011.
[63] GAVI Alliance Board Meeting, 16 - 17 November 2011, Dhaka, Bangladesh, Final Minutes.
[64] GAVI Alliance Board Meeting, 16 - 17 November 2011, Dhaka, Bangladesh, Final Minutes.
[65] Chee et al. 2007.
[66] http://globalhealthsciences.ucsf.edu/pub/health-systems-funding-platform%E2%80%94-primer
[67] https://globalizationandhealth.biomedcentral.com/articles/10.1186/1744-8603-9-9

***Key changes to the PBF model after Board-approval in November 2011***

Significant changes were made to the PBF model after the November 2011 Board meeting in terms of country groups and the split of payments into core funding and performance payment, following a hand-over from the Secretariat's Policy and Performance Team to the Country Programmes team.

Key informants reported that, in the development of the model that was presented to the Board in November 2011, there was limited consultation of country stakeholders. According to the key informants, Rwanda was the only country where the model was presented in-person.[68] In addition, a few other countries were consulted through phone conversations. However, as one Secretariat member reported, the country consultation process was not "meaningful enough." Gavi staff also reported that there was only little ownership within the Country Programmes team.

After the hand-over within the Secretariat, the responsible Gavi staff members from the Country Programmes team consulted with Gavi's country managers who – to a certain extent – also discussed the new PBF model with country representatives. In addition, the model was presented to regional representatives from WHO and UNICEF. One outcome of these conversations was that the model was considered to be too complicated for countries. There were also concerns that countries could lose substantial HSS funding because they would not be able to qualify for performance payments.

Following these consultations, the PBF model was substantially changed and simplified. The current PBF model, described in the introduction of this review (Figure 1 and Table 1), was introduced. This model significantly differs from the originally Board-approved model. It no longer includes three country groups but two (so there is no specific country group with coverage below 70%). In addition, the share of the performance payment is the same for all countries (i.e. the proportion of the amount at risk did not end up being linked to baseline coverage levels). In addition, the amount at risk was lower than in the previous model – as such, the eligibility criteria for the performance payment were substantially changed.

The handover within the Secretariat resulted in delays in the roll-out of the PBF model. The work was overseen by a new Technical Advisory Group on Health System Strengthening. Different revisions of the PBF model were presented to the Executive Committee, which initially did not accept the changes to the Board-approved version but eventually approved the revised PBF model.[69] In parallel, Gavi's HSIS team was working on the operationalization of the model. This included the development of operational guidelines, which clearly laid out the timing and frequency of payments, clearance and approval, guidance on verification and data, among other topics.[70] The first assessments of eligibility took place in 2014 for five countries (see Section 4.3 for more details). Additional changes were made to the model after it was rolled out in 2014.[71]

[68] A representative of Nigeria was on the PBF-TT. The model was also discussed at the PPC, which also includes representatives from countries.
[69] Interview with previous Gavi staff.
[70] Gavi Alliance, the Vaccine Alliance: Operational Guideline: 3.15 Performance Based Funding (PBF).
[71] An internal Gavi PBF Operation Guideline was issued which noted that countries with at least 6 months of implementation in its first year of HSS grant, could be considered for PBF eligibility the following year. Under the new HSIS

## 4.2 To what degree is Gavi's PBF model "fit for purpose"?

**This section addresses the following questions:**

▪ Are the implicit and explicit assumptions – the underlying theory of change – underpinning the PBF model robust and appropriate? Is the design of the current PBF model appropriate and sufficient to lead to improved immunization outcomes and equity?

- Are the assumptions on reaching the performance indicators realistic? Are the goals achievable in the suggested timeframe, or are they overly ambitious?
- Is country performance accurately be measured and verified? Is the measurement approach sound? What are the cost implications of the measurement?
- Is the size of the performance payment significant enough to incentivize countries to give more attention to immunization coverage and equity?
- Is the timing of the performance payment closely linked to the performance itself? Is the payment received quickly after the implementation?
- To what extent does the PBF model consider the specific challenges faced by countries with very low capacity, and those affected by conflict and humanitarian emergencies?

▪ Is Gavi's PBF embedded in a learning environment?

**Key findings: Gavi's PBF model suffers from several design challenges. Increases in coverage are difficult to achieve within one year and, even more importantly, current measurement systems are not precise enough to measure such changes (which was already recognized by the PPC in 2011). There is a long delay between the successful implementation that triggers a reward, and the actual receipt of the reward. PBF should always be accompanied by evaluation and learning—yet we found that Gavi's PBF was not embedded in a learning environment. On a positive note, country flexibility in how the reward is used is a valuable design feature.**

Below we review the design of Gavi's PBF model based on the following criteria: (i) achievability of performance targets; (ii) robustness of measurement and verification approach; (iii) size of the performance payment; (iv) speed of performance reward, (v) equity across countries, (vi) flexible use of payments; and (vii) being embedded in a learning environment.

### *i) Are the assumptions on reaching the performance indicators realistic? Are the goals achievable in the suggested timeframe, or are they overly ambitious?*

Achieving increases in immunization coverage within one year is very difficult. A range of key informants argued that increase in coverage usually require more time, and that the underlying theory of change is thus unrealistic and flawed. For example, one SCM argued:

framework, which entered into force in January 2017, the PBF approach was modified. As discussed in section 1.2, there was a move from a model where the performance payment is integrated into a country's ceiling allocation to a supplemental model.

> "We should more look into outputs and some of the processes. Coverage does not change that much in one year. If you look at the real number of immunized children – that is fine – or use certain activities and outputs from the HSS grant performance framework."

As highlighted in Section 3, the SMI measures performance after 18-24 months, and even high capacity countries can find it hard to achieve the performance targets within such a short time frame. For the low-coverage group in Gavi's PBF, the theory of change is indeed very ambitious (for the high-coverage group, only maintaining coverage levels is required).

In addition, outcome indicators are sensitive to a range of co-founding factors. There are many factors that affect coverage that are beyond the control of governments. For example, in Haiti, a maternal and child health project that was funded by a bilateral donor came to end about two years ago. This project covered significant funding for immunization, which led to EPI funding gaps and contributed to declining coverage. As mentioned in the rapid literature review, a host of contextual factors can influence the success of PBF schemes.[72]

***ii) Is country performance accurately be measured and verified? Is there robust data to assess performance? What are the cost implications of the measurement?***

The PPC report on PBF, discussed at the September 2011 PPC meeting, highlighted measurement and data quality challenges: "available measures of immunisation coverage are not sufficiently precise to determine whether a reported one percentage point increase in coverage is a true increase in coverage as opposed to measurement error" (Box 3).

In November 2011, the Board, in the context of its PBF discussion, noted that "data quality is a matter of concern (…) and recognised that more work is needed to assess and improve country data systems and to advance innovation in coverage estimation".

Key informants interviewed for this review reported that participants at the November 2011 Board meeting raised substantial concerns about measurement. They argued that the PBF model is too sensitive to measurement errors – as small increases in coverage (one percentage point) are extremely hard to measure in a reliable manner. So, as early as 2011, there were already substantial doubts that eligibility for reward payments can be measured accurately.

A range of key informants also referred to the measurement challenge, with one key informant even mentioning that it is "impossible to track one percentage point increases in coverage." Key informants highlighted that there is substantial uncertainty around both WUENIC and country administration data. Until 2015, countries had to use their own administrative data by default, which made it difficult for them to earn performance payments. The difference between WUENIC and country coverage data is up to 20-25 percentage points for certain countries. As we discuss in Section 6 below, this major discrepancy is an inherent weakness of the PBF model as the verification mechanism prevented countries with weak data from accessing performance rewards.

[72] Renmans D, et al. Opening the 'black box' of performance-based financing in low- and lower middle-income countries: a review of the literature. *Health Policy and Planning* 2016;31: 1297-1309.

In addition to these problems with the verification, SCMs reported other sensitivities related to the indicators. For example, SCMs pointed to instances where countries introduced new census data, which led to a boost in coverage (e.g. in Honduras). For equity, another challenge relates to very small birth cohorts in small municipalities (i.e. a very small denominator), which can also quickly result in the fact that countries do not reach the equity indicator.

On a positive note, the verification approach is low cost, especially compared with verification methods used by the SMI and the World Bank, and is relatively simple.

| **Box 3: 2011 PPC report on PBF – measurement challenges** |
| --- |
| "Under the recommended design (…), countries must increase coverage in order to receive performance payments, so countries with stagnating or decreasing coverage would not receive performance payments. Countries that increase coverage by a single percentage point would, however, qualify for a performance payment, even though available measures of immunisation coverage are not sufficiently precise to determine whether a reported one percentage point increase in coverage is a true increase in coverage as opposed to measurement error. This could pose a significant reputational risk for the GAVI Alliance. This risk is, however, small in monetary terms since the size of performance payments is proportional to increases in coverage and children immunised—small increases in coverage and children immunised would therefore lead to small performance payments. This risk could be further mitigated by introducing additional checks and balances that use household survey data over longer time periods to assess whether the reported increase in coverage over the time period corresponding to the cohorts measured by the different surveys is supported by the independent survey data."[73] |

### *iii) Is the size of the performance payment significant enough to incentivize countries to give more attention to immunization coverage and equity?*

Key informants also questioned whether the incentive is large enough to change the behavior of countries, or whether the share of the incentive funding needs to be larger (yet others raised strong concerns about predictability if the core tranche were to be smaller).

The ISS review suggested that the incentive payment of US$ 20 per additional child is too small, and this concern was also reflected in the PPC PBF paper from September 2010. This paper states that "Under the ISS window, countries can receive reward payments— albeit very small payments—for vaccinating a single additional child, even if coverage is stagnant or decreasing."

Compared to ISS, the performance payment per additional child immunized increased from US$ 20 to US$ 30 for the group with coverage below 90%. However, key informants argued that even $30 per child is too small– especially compared to other investments made by Gavi and the support from other bilateral and multilateral donors. Overall, key informants also indicated that the overall amounts of the performance payments are too small to incentivize better national immunization planning and implementation. For example:

[73] Gavi Alliance: Report to the Programme and Policy Committee, 28-30 September 2011.

> "A lot is not performance oriented. Not PBF. You will still get your programming component. The top up is performance based. But not the whole model."

> "Overall, the amount of PBF money provided is too small to serve as an incentive for many countries."

Key informants also reported that the change in the PBF model after the 2011 Board meeting watered down the incentive mechanism:

> "It became less of an incentive-based scheme – the incentive was kind of moved. It was better to convey it to countries."

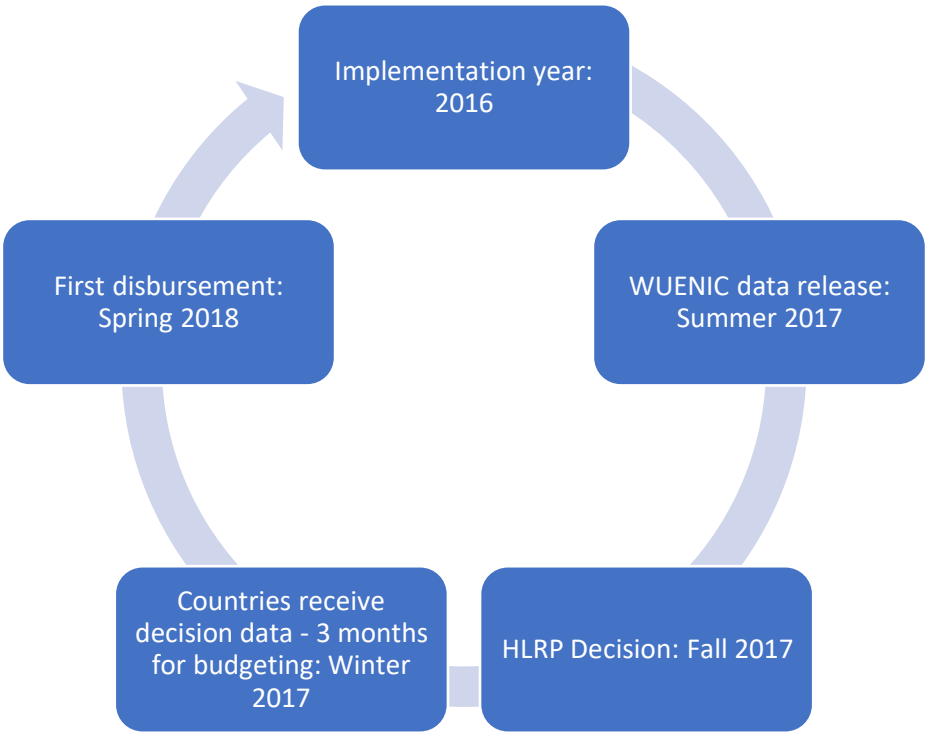
In Section 6 below, we assess further to what extent the model serves as an incentive. Overall, findings indicate that some countries were motivated to improve their EPI, and, for example, changed their strategic planning. At the same time, other countries were not effectively incentivized. However, to further assess the magnitude of the potential size of PBF rewards, we developed a model for countries with DTP coverage below 90%. The potential awards were calculated as the gap between current coverage and 100% coverage multiplied by US$30. The data shows that with the current incentive structure of US$30 bonus per additional immunized child, low DTP3 coverage countries had the potential to earn between US$71,342 and US$29 million (mean: US$4.9 million) per year in DTP3 payments and between US$79,269 and US$36 million (mean: US$5.3 million) per year in MCV1 payments. This shows that the financial incentive is not necessarily small but depends on the specific coverage level and the size of the birth cohort.

#### *iv) Is the timing of the performance payment closely linked to the performance itself? Is the payment received quickly after the implementation?*

Gavi's PBF approach requires the availability of actual WUENIC data to verify the results reported by countries based on their own administrative systems. As WUENIC data are released in mid-July each year, PBF eligibility decisions for the implementation year can usually only be made by the HLRP and the MDs in October (or later). Once the decision is taken, countries have three months to prepare PBF budgets and workplans (Figure 3).

For example, if the implementation year is 2017, the verification takes place in July 2018. The HLRP can then decide about eligibility in October 2018. In an ideal scenario, countries would then prepare a workplan and budget to receive the performance payment in April 2019 – 16 months after the end of the implementation year. However, often it takes longer as the PBF disbursement data indicates: as of December 2017, only **US$15.6 million** of the funding has been disbursed.

**Figure 3: PBF process: From implementation to disbursement of funding (illustrative)**




The benefit of a strong performance is only felt substantially later. The payment is **de-linked** from the successful implementation that triggered it initially. This differs from other performance- and results-based schemes, like the World Bank's scheme, which provide the reward more quickly (in many countries, the World Bank usually rewards health facilities on a quarterly basis for good performance). Key informants considered this to be another design error. One of the contextual factors that is believed to be associated with improved outcomes is the timeliness and predictability of the performance payment.[74]

One suggestion made by key informants was to find a lighter decision-making process. Key informants questioned whether the HLRP needs to decide on performance payments given how the current model works. In the past, the HLRP may have imposed conditions on the payment (e.g. audit reports) but it never overturned eligibility.

***v) To what extent does the PBF model consider the specific challenges faced by countries with very low capacity, and those affected by conflict and humanitarian emergencies?***

More specifically for Gavi's PBF model, the PBF indicators for countries with baseline coverage below 90% are more ambitious compared to countries with baseline coverage at or above 90%. The group with lower baseline coverage levels needs to increase DTP3 and MCV1 coverage to gain a performance payment, while the higher baseline coverage group needs to only maintain coverage level. As reported by Gavi staff, keeping coverage at high levels (above 90%) is easier than reaching such levels. Since

[74] Witter S, et al, Performance-based financing as a health system reform: mapping the key dimensions for monitoring and evaluation. *BMC Health Serv Res* 2013;13:367.

Gavi introduced its model, only one country fell below 90% for one year (out of 16 countries that received PBF awards in the category >90% coverage).

In this context, Gavi's model does not differentiate between more stable but underperforming countries and countries in emergencies. Currently all countries with a coverage rate below 90% are together in one group – countries like South Sudan with DTP3 coverage of 30% and countries with coverage of 80%. And if countries fall below their baseline, as Haiti did, it will become impossible to earn a performance payment.

In addition, coverage targets might also be a difficult measure because of population growth. A large country like Ethiopia may vaccinate many additional new children but will still not achieve increases in coverage because of an increase in population growth of 2.5% per year (if growth continues, Ethiopia will have a population of 190 million people in 2050).

#### *vi) Flexible use of payments*

The guidance for the use of PBF funds was narrowed down from the "country's health sector" to "immunization-related activities" exclusively. The initial model did not attach conditions to the use of performance payments (except that the payment could not be used for Gavi co-financing). However, while there is a move towards a more directed use of the HSS funding, country representatives reported that they were able to use the PBF payments in a flexible manner. As such, Gavi's PBF still meets the requirement in terms of flexible use.

#### *vii) Is Gavi's PBF embedded in a learning environment?*

PBF was initially seen as a learning exercise but at best very little learning took place. For example, Gavi considered introducing an equity component for all countries if the experience with high-coverage countries turned out to be positive. Gavi also considered further improving data systems and seeing if the chosen verification mechanisms worked for countries. However, there is no discussion of PBF at Board level and reporting currently only takes place at PPC level (but data is only provided in an annex to the main report). In fact, one Gavi Secretariat member described PBF as a "mysterious objective in Gavi, of which many do not know about." In fact, Gavi never laid out the conditions for learning because, as we discuss in Section 5, the tools for monitoring the results of PBF were never developed.

## 5. Implementation of the PBF model

**Section 5 addresses the following questions:**

- How well has the PBF model been communicated at global and country level? Is there sufficient understanding of it at both levels?
- What have been the main challenges to the implementation of the PBF component in terms of budgeting, work planning and implementation at global and country level? How much alignment is there with country planning and review processes?
- To what extent is there (financial) reporting on the use of the performance payments, and PBF-related results? Is M&E conducted as an integral part of a country system?
- Is Gavi's PBF harmonized with other result-based financing schemes at country level?
- Were there specific issues related to the implementation of HSS grants overall?
- To what extent did changes to the PBF design affect the relevance of the PBF to meet its intended objectives?

### 5.1 How well has the PBF model been communicated at global and country level? Is there sufficient understanding of it at both levels?

**Key findings: There is poor understanding of Gavi's PBF model at country level. In particular, countries that never qualified for a performance payment have very little knowledge of the PBF scheme. This lack of understanding impedes effective implementation of the PBF model.**

Clear communication on the PBF model is a critical precondition to ensure its effective implementation. Countries need to fully understand how to qualify for performance payments, as they otherwise will not be able to undertake the required action for earning the performance reward.

Key informant interviews with country representatives show that knowledge about Gavi's PBF model at country level varies substantially. While certain countries are aware of the details of the PBF model, others lack a basic understanding of the main parameters (some countries were not even aware of the existence of the model).

Country representatives with good knowledge of the performance indicators and the PBF process tend to be from countries that qualified for the performance payments in the past (e.g. Tanzania, Honduras, and DPR Korea). At the same time, there are still several countries, where Gavi's PBF is not well understood (e.g. Haiti, Solomon Islands). For example, country representatives broadly referred to the indicators of the grant performance framework. Others referred to the number of additional children immunized, without being aware of the need to increase coverage. Yet others were unaware of the annual PBF assessment process. In one country, one key informant reported that she consulted multiple people ahead of the interview to find out whether the country ever qualified for Gavi's PBF but nobody was able to answer this question.

Our key informant interviews with UN country offices (WHO, UNICEF) and other technical partners (e.g. CHAI) show that these country partners have a better understanding of the PBF model. They also reported that they explained the model to EPI managers and other government staff who previously lacked basic knowledge on the model.

For several countries, SCMs indicated during the review process that it would be of little value for the evaluation team to conduct interviews with countries that did not qualify, as the country representatives would not know much or anything about Gavi's PBF model. As countries never qualified for PBF payments, they are not familiar with the model. These countries were also very much occupied with the implementation of the general HSS grant, and thus lack capacity to qualify for a PBF reward. Staff turnover was given as another major reason for the lack of knowledge on the PBF model.

Among SCMs and POs, knowledge of the PBF model was strong overall, although to some extent also variable when it came to the details of the PBF model. Frequent changes to the PBF model appear to have contributed to this variability.

SCMs highlighted that there is very little internal guidance on Gavi's PBF model, and that there is also little guidance on PBF that they could share with countries, especially compared to other Gavi policies and types of support. The operational guidelines and the HSIS Framework provide a few details, but overall the guidance is rather limited.[75]

This lack of guidance is also reflected in the following interview statements:

> "While Gavi HSS is well understood, there needs to be more explanation and information on PBF. The PBF guidelines are tough to understand and not practical for the country" (UN technical partner).
>
> "We need more guidance, and there is not a lot of awareness and understanding of PBF at country [level]. We do not use the alliance partners to reiterate e.g. the strategic use" (Gavi SCM).
>
> "We make it up as we go along. There are no systems and processes" (Gavi SCM).

During the early implementation of the PBF model, there were also a number of early communication challenges. For example, in 2013/14, decision letters for 14 countries had to be revised because the submission process overlapped with the development of new PBF timelines. The 14 countries did not know at the time of submission that 20% of the HSS funding from Year 2 onwards would be based on performance. The countries thus had to change their budgets and M&E frameworks, including for the six mandatory outcome indicators. The HSIS provided support to countries during this process.

Overall, the guidance and communication on PBF affects the implementation of the PBF model. In this context, a meta-evaluation of Gavi's HSS found that "Gavi's model of delivery for HSS in terms of

[75] Operational Guideline: 3.15 Performance Based Funding (PBF).

guidance and support from the Secretariat and Partners has not functioned effectively across the grant cycle, with most evaluation reports recommending more active guidance and communications from Gavi through a more "hands-on" model."[76] Gavi has moved towards this hands-on model – there is more guidance on HSS and the Country Programmes team has significantly grown in recent years. At the same time, PBF, as a critical component of the HSS, did not receive the same attention.

## 5.2 What have been the main challenges to the implementation of the PBF component in terms of budgeting, work planning and implementation at global and country level? How much alignment is there with country planning and review processes?

**Key findings: The alignment of Gavi's PBF with country processes is limited. The PBF model also involves additional transaction costs at global and country level. However, compared with other HSS grant processes, countries find the PBF approach to be less burdensome.**

The verification and approval process results in a lack of alignment with country planning processes. For example, in Tanzania, planning for the upcoming year usually ends in June. The decision on the PBF reward is only made later in the year, and the payment would only arrive in the subsequent year. This makes planning more complicated and to some extent also unpredictable. The Gavi guidelines assume that "Countries (…) update their existing health system strengthening grant budget with the additional performance payment amount, in order to ensure integration of additional activities with existing support, reduce potential duplication and facilitate annual reporting."

One SCM reported:

> "We are not aligned with country processes. Two design challenges really affect the lack of alignment – we are depending on data and our HLRP. We cannot do anything until we have WHO/UNICEF data in July – so we lose half a year. Then the busiest HLRP panel is in October and that usually kicks off the process. Countries are already implementing the next year of HSS funding, while we are running behind schedule. From that point we asked what they will invest the PBF in. You usually disburse in March, April, and that is the quickest we ever disbursed. And these are good performing countries."

A country level stakeholder commented in a similar way:

> "The PBF planning process is not well synchronized with local planning process. EPI programming is following Gavi's agenda almost completely, which is a good and rational planning approach. But the counterpart is that tools and schedules are not well adapted— sometimes it requires additional work for country staff to be able to align with Gavi requirements. It may be useful for the country to be able to show what is on their schedule and what is their planning process. We face weaknesses in planning, reporting, M&E. To

[76] Gavi, the Vaccine Alliance 2016: Meta-Review of Country Evaluations of Gavi's Health System Strengthening Support. Prepared by CEPA.

apply Gavi's guidelines is hard. We need a framework to follow. There's no opportunity for the country to set its own pace or follow its own agenda."

As reported by Gavi staff, processes may take longer if alliance partners are responsible for the financial management at country level. This happens particularly in fragile countries, with weaker capacity to guarantee financial oversight. In these cases, the ministry of health needs to request the funding directly from the technical partners to implement it. This process also adds to the transaction costs.

In addition to the lack of alignment, there are also **high transaction costs** involved (for, at times, small amounts) at global and country level due to parallel processes. The PBF model requires the creation of one-year budgets and workplans, as well as one-year grant agreements. During the key informant interviews, country representatives tended to be less concerned about these processes than Secretariat staff, which found the process to be very transaction-cost heavy. Countries in fact reported that they find the PBF process rather light compared to Gavi's general HSS grant processes.

### 5.3. To what extent is there (financial) reporting on the use of the performance payments, and PBF-related results? Is M&E conducted as an integral part of a country system?

**Key findings: There is very little evidence on and monitoring of PBF both at global and country levels. This evidence gap makes it difficult to measure the effectiveness of the PBF model and impedes learning.**

There is very little specific information on PBF in country proposals and JRFs. Country proposals do not include any information on PBF. Countries only need to make clear which data for the verification they wish to use.

Gavi collects very little information specific to its PBF scheme. Through the joint appraisals, countries annually report on *overall progress* of Gavi's vaccine and cash grant support, but the joint appraisal templates do not request specific PBF information. Results are usually merged together with the overall results of Gavi's HSS support. Similarly, the financial reporting usually does not separate between general activities under the HSS grant and those funded by performance payments.

When it comes to JRFs, all information is usually merged with general information on the HSS grants, with only few countries having specific performance data for PBF. This is usually also the case for the financial reporting – financial reports usually do not differentiate between expenditures for core HSS and the performance payments. There are no PBF reporting templates, and there is no guidance on the frequency and type of reporting for PBF. As a member of the Secretariat reported:

> "It comes more down to the SCM. He/she may want a specific metric. No template for reporting on PBF. We are not clear on reporting frequency, type of reporting."

A budget for PBF was recently introduced but budget templates for HSS and PBF HSS look very different, and some SCMs indicated that there might be easy options for a consolidated version:

> "There is a new budget template – current active grants in the same file. Not really handy. Very big file."

In terms of country ownership and in-country oversight, key informants raised concerns that ICCs and HSCs are less involved in PBF compared to other types of support from Gavi. Key informants provided examples of countries where the EPI manager did not ask for ICC endorsement of PBF budgets. Overall, there are indications that the EPI (or HSS) managers are in more control of the PBF compared to other Gavi support.

> "EPI managers are not always informing the other stakeholders in countries. Wider stakeholder participation does not apply to PBF planning and budgeting - sometimes there is information to ICC but not like for other Gavi support" (Gavi HSIS).

Interviewed country representatives, however, made clear that the PBF is discussed by the ICC – so the extent to which PBF complies with the wider principle of country ownership varies across countries.

### 5.4. Is Gavi's PBF harmonized with other result-based financing schemes at country level?

**Key findings: Gavi's PBF is not integrated with other RBF schemes at country level. As new RBF mechanisms are currently being established in many countries, going forward, it will become increasingly important to ensure better harmonization.**

Gavi's PBF is focused on the national level rather than rewarding districts, health facilities, or health workers. There is thus relatively little integration with other results-based-financing schemes. For example, there is no integration in Rwanda, where a donor-led RBF mechanism exists. In Nicaragua, there are also no links with the World Bank's RBF program and this is also true for other countries like Kenya and Tanzania. In DPRK, there is also a lack of a joint coordinating mechanism, so there are no synergies between Gavi's PBF and that of the Global Fund. In Bangladesh there is a pooled funding mechanism but there are also no linkages between PBF schemes.

Given how Gavi's PBF model works, it is not much of a surprise that there are few linkages. Unlike other PBF/RBF schemes, Gavi's model may not necessarily involve payments to health workers or facilities at subnational level. It is a government-focused mechanism, which provides additional funds to the MoH directly rather than providing incentives to subnational entities (or individuals).

However, going forward, the integration with other PBF or RBF schemes will become more important. For example, the GFF is introducing PBF programs in Lao PDR and Mozambique. Experience from current countries indicate that it will be a challenge for the countries to integrate Gavi's model with other funding schemes; this concern was also raised by interviewed country level stakeholders.

Integration matters because there is evidence that having simultaneous, non-aligned schemes can reduce the effectiveness of PBF, especially if the schemes have competing priorities.[77]

[77] Renmans D, et al. Opening the 'black box' of performance-based financing in low- and lower middle-income countries: a review of the literature. *Health Policy and Planning* 2016;31: 1297-1309.

## 5.5. Are there specific issues related to the implementation of HSS grants overall?

**Key findings: Progress in the implementation of the HSS grant does not affect the decision about PBF eligibility. Countries may struggle to effectively use the additional funds in light of already existing absorption issues that many countries are facing.**

Progress in the implementation of the HSS grant does not affect the decision about PBF eligibility. For example, countries can have low use of their HSS grants but would still earn additional funds through the PBF model. Countries may struggle to effectively use the additional funds in light of already existing absorption issues – this is, for example, the case in countries like Ghana, Tanzania, Timor Leste, and Mozambique.[78] As such, we find that these absorption issues are a challenge for many countries and that they are unrelated to population size.

Country representatives also argued that the relevance of the PBF model is diminished because of the general absorption issues:

> "I trust in the PBF approach but there are two limitations here. First, data quality and availability here are poor. It is hard to measure performance. Second, country capacity to implement and use funds is limited. We face a real administrative and financial limitation in processes making it hard to use funds. In the 4th year of HSS implementation, we have only used around 25-50% of all funds available for the 5 year program. There's been a huge delay in fund implementation. So PBF is welcome but *so far the country is struggling to even implement the whole Gavi HSS grant*. So, overall, data limitations and weak country capacity make the PBF less relevant to us. So PBF really isn't a big incentive."

Limited absorption capacity is also concern for other countries, which could potentially receive very high amounts of PBF. Pakistan's overall HSS grants amount to US$100 million, and due to the fact that Pakistan uses the survey mechanism and (according to current estimates) increased coverage significantly, it may receive a performance payment of up to US$50 million. While there is a general demand to invest in immunization (e.g. in urban slums), there are doubts that Pakistan could effectively use such a large amount if the burn rate of the current grant is considered.

In addition, Gavi's HSS support has been evolving. The guidance has been focused on vaccination outcomes and equity rather than HSS more broadly. Gavi is thus moving to a more directed model (similar to the Global Fund model). Key informants argued that it would thus be useful to continue with a more directed and focused PBF approach – one that better guides countries on how to use the PBF support (rather than a "letting a hundred flowers bloom" approach). From this angle, it would also be useful to track PBF separately from the general HSS support.

[78] Previous HSS evaluation have also shown that Gavi HSS grants experienced substantial delays in implementation; see e.g. CEPA'S Meta-Review.

## 5.6. How did countries use the performance payments? To what extent did the use of payments affect the performance?

**Key findings: Countries mostly used the PBF funding to fill gaps in their budgets. More recently, countries also used it to co-finance the Cold Chain Equipment Optimisation Platform.**

Countries mostly used the PBF funding to fill gaps in their budgets – such as salaries, bonuses, supervision, warehouses, vehicles, and surveillance (see Box 4 for an example). This wide range of uses of the performance payment reflects the fact that Gavi has not provided concrete guidance to countries on how to use it, again indicating a lack of structure and guidance around the PBF model.

> "The idea was to give more flexible funds for countries - lighter touch and more creative and innovative but this does not take place" (SCM).
>
> "Countries do more of the same – they could use the money in a better way" (HSIS team).

| **Box 4. Use of PBF payments in DPRK** |
|---|
| *1: Refurbishment of EPI clinics*: DPRK has selected 300-400 clinics for refurbishment, and the PBF payments have been very useful for this purpose. The last external review, coordinated by WHO, which took place after pentavalent vaccine introduction, concluded that EPI clinics needed refurbishment (the poor conditions were keeping patients away, and the freezing temperatures were causing vaccines to get frozen). Refurbishment has included charcoal-based heaters, blood pressure machines, and stethoscopes. |
| 2: *ICT equipment*. DPRK only has electronic records from the central to provincial levels, not below that to counties or sub-counties (here the flow of information remains paper-based). The PBF has been used for computerization—starting with a small rollout, which will then go national. The plan is for each of three key entities at county level—the county medical warehouse that stores vaccines; the immunization unit; and the hygienic & anti-epidemic station—to have at least one computer. |
| *-3*: *logistical management information software system*—the aim is for all warehouses to connect with each other. |
| *-4: Quality improvement*. The review conducted after introduction of new vaccines also noted that while DPRK has high coverage and high equity, there is now a need to shift to *quality* of services, specifically (a) improved M&E, and (b) improved supportive supervision. The PBF reward purchased two vehicles that allow the central supervisory team to go to counties to do supportive supervision. PBF has also been used to support printing of a supervisory checklist, per diems for the supportive supervision, and health worker training. |

Many countries, like Tanzania, Lao, and Mozambique, also use the PBF reward to co-finance the Gavi's **Cold Chain Equipment Optimisation Platform (CCEOP)**, which is helping countries modernize

cold chains with high-performing equipment. In December 2016, in light of very strong demand from countries, the Gavi Board approved a formula for calculating a budget ceiling for this support which amounts to US$250 million for all eligible countries. It requires a co-investment from countries and many countries use the PBF to pay for it.

As we discuss in Section 6, at times countries invest in a more strategic way - for instance to improve equity (for example, in Lao PDR, PBF payments were channeled to districts with lower coverage).

## 5.7 To what extent did the changes to the PBF design affect the relevance of the PBF to meet its intended objectives?

**Key findings: There is no evidence that the changes that were made to PBF model after 2014 affected its implementation. The most impactful change might be the introduction of new measurement options in 2015 but its impact has not yet been evaluated.**

The most substantial changes to the PBF model were made before its roll-out in 2014. The largest changes made after 2014 were (i) the narrower use of the PBF funding (most recently defined in the HSIS Framework), (ii) the change to the supplementary model, and (iii) the introduction of new measurement options in addition to the default option (Box 1).

With respect to the narrower use of PBF funding, country representatives reported that they find the flexibility of the performance payments one of the key strengths of the PBF model. There were no concerns that the narrower focus might negatively affect the PBF in the future. Overall, we thus find no evidence that the narrower definition on use of PBF support has a negative impact. However, many countries were not well aware of the details of the policy, and this to some extent impacts on the robustness of this finding.

One immediate impact of the change to the supplementary model is that countries have more funding predictability when they develop their PBF budget (for high coverage countries, a smaller funding share is at risk; see Box 1). However, countries still need to develop additional budgets for the PBF as discussed above. Overall, this change is rather small and does not fundamentally change the PBF model.

New measurement mechanisms were introduced in 2015, including the option for countries to to use the WUENIC data for their baselines and the annual PBF reviews, but these are not widely used yet (Box 1). Only two countries covered by the timeframe of this review requested an alternative data option for measuring 2016 performance. Going forward, a number of countries with new HSS grants opted to use the WUENCE, so in 2019, their performance will be assessed based on the WUENIC option for the first time. Given that this is potentially a very important change to the model, we developed a few hypothetical models to assess the likely impact of this change in the future (see Section 8).

# 6. Results of Gavi's PBF model: Contributions to coverage and equity

**Section 6 addresses the following questions:**

- Has the PBF model provided equal access to reward payments across countries?
- Did the PBF approach incentivize more action towards improved immunization coverage and equity at country level? If so, what kind of action?
- To what extent did Gavi's PBF contribute to improvements in EPI coverage at country level? What were the major factors influencing the achievement of these results?
- What have been the unanticipated consequences of PBF at country and global levels (if any)?

## 6.1 Access to Gavi PBF: Which countries benefit, which do not?

In this section, we address the question of whether the PBF model has shown equity *between* countries. In other words, has the model provided equal access to reward payments across countries? To answer this question, we examined data on which countries were assessed for eligibility, which were found to be eligible, and how this eligibility varied by baseline DTP3 coverage.

**Countries receiving performance payments**

Over the period 2014-2017, 69 countries were assessed for eligibility (5 in 2014, 15 in 2015, 21 in 2016, and 28 in 2017). As shown in Figures 4 and 5, performance payments went to two countries in 2014, 8 countries in 2015, 9 countries in 2016, and 13 countries in 2017. The total amount of the performance payment was US$3.2 million in 2014, US$6.5 million in 2015, US$8.3 million in 2016, and US$16.7 million in 2017 (a total of US$34.8 million over the 2014-2017 period).

Country performance payments were allocated 32 times over the four-year period, with many countries being eligible for performance payments in multiple years. A total of **16 different countries earned a reward** in at least one of these four years (Table 7). Ten out of these 16 countries received awards in multiple years (Burundi, Lao PDR, Nicaragua, Rwanda, Sudan, Tanzania, Zimbabwe, Honduras, Korea, and Solomon Islands).

**Figure 4. Number of countries assessed and eligible for performance payment, 2014-2017**

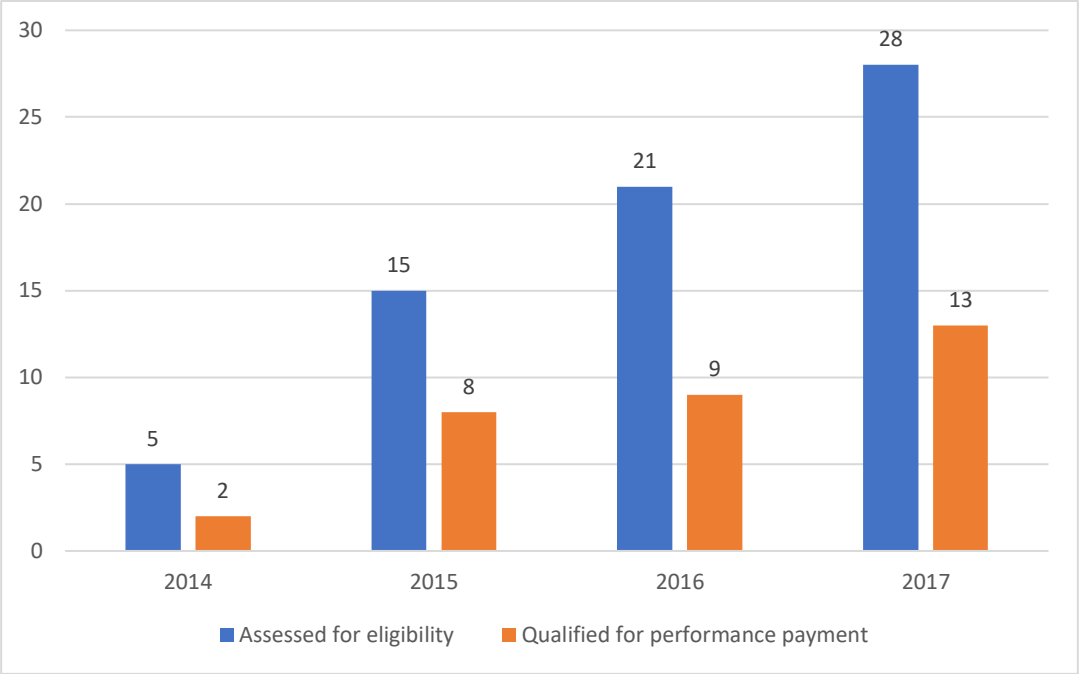



**Figure 5. Number of countries eligible to receive performance payment and total amount allocated, 2014-2017**

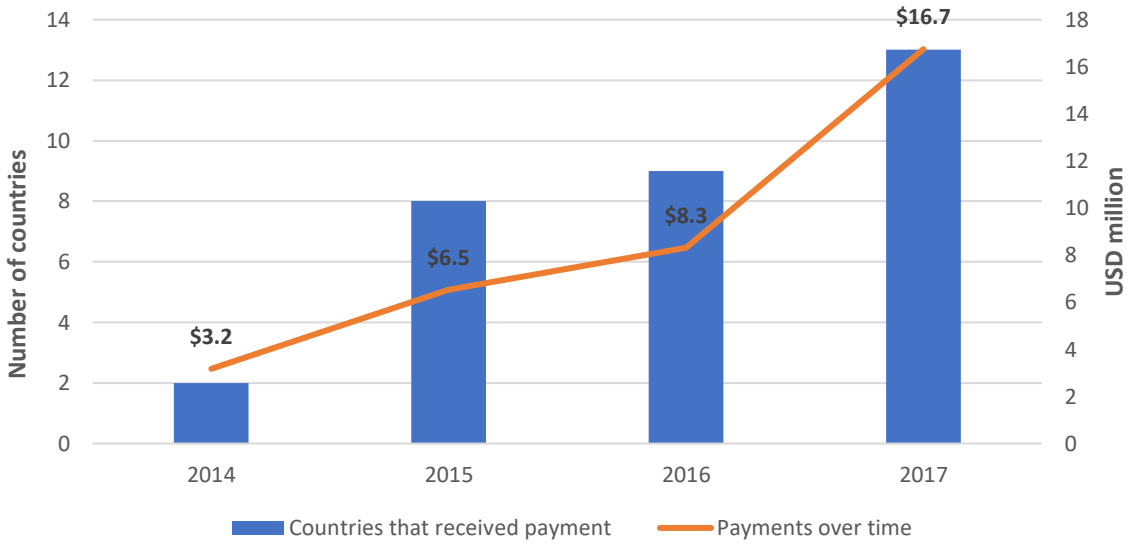


**Table 7. Number of countries assessed for PBF eligibility, countries allocated performance payments, and size of payments, 2014-2017**

| Year | No. of countries assessed for PBF eligibility | Countries allocated performance payments | Total performance payment allocation to the country (in US$) |
|---|---|---|---|
| 2014 | 5 | Burundi | 1.720.000 |
| | | Lao PDR | 1.454.370 |
| | | | **3.174.370** |
| 2015 | 15 | Burundi | 1.720.000 |
| | | Lao PDR | 441.540 |
| | | Nicaragua | 240.000 |
| | | Rwanda | 984.000 |
| | | Sudan | 1.584.000 |
| | | Tanzania | 800.000 |
| | | Timor Leste | 73.020 |
| | | Zimbabwe | 680.000 |
| | | | **6.522.560** |
| 2016 | 21 | Lao PDR | 393.570 |
| | | Honduras | 44.280 |
| | | Tanzania | 1.600.000 |
| | | Korea | 2.622.400 |
| | | Sudan | 1.584.000 |
| | | Burundi | 860.000 |
| | | Nicaragua | 120.000 |
| | | Rwanda | 984.000 |
| | | Solomon Islands | 120.000 |
| | | | **8.328.250** |
| 2017 | 28 | Bangladesh | 5.228.000 |
| | | Djibouti | 36.480 |
| | | Ghana | 860.000 |
| | | Honduras | 1.200.000 |
| | | Korea | 2.622.400 |
| | | Mozambique | 1.489.620 |
| | | Nicaragua | 240.000 |
| | | Rwanda | 984.000 |
| | | Senegal | 443.130 |
| | | Solomon Islands | 120.000 |
| | | Sudan | 1.584.000 |
| | | Tanzania | 1.600.000 |
| | | Zimbabwe | 340.000 |
| | | | **16.747.630** |
| **Total** | **69** | | **34.772.810** |

**Breakdown of reward payments by country coverage group**

**Key findings: Gavi's PBF model mostly benefits countries with high (≥90% DTP3) coverage, while countries with low coverage (<90% DTP3) find it difficult to earn reward payments due to data issues and/or lack of progress.**

A striking finding of our analysis is that of the 16 countries that received a reward payment, **11 had a baseline DTP3 coverage of at least 90% and only five had coverage below 90%** (Table 8). Looking at the payment amounts, the 11 countries with high coverage at baseline accounted for 87.5% of total allocated payments (US$30.4 million), while the five countries with low coverage at baseline received only 12.5% of total payments (US$4.3 million) (Table 8). A similar pattern was seen with disbursements: high coverage countries received 80% of disbursements (US$12.5 million) and low coverage countries only 20% (US$3.1 million).

Out of the total PBF payments over 2014-2017, **the top five recipients of allocated payments accounted for 68% of all payments**—these five countries are all in the high baseline coverage group (≥90%) and include Bangladesh (US$5.2 million), Burundi (US$4.3 million), Democratic People's Republic of Korea (US$5.2 million), Sudan (US$4.75 million), and Tanzania (US$ 4 million).

**Table 8. Performance payments and disbursements broken down by baseline coverage level (as of end 2017)**

| | Countries with DTP3 Coverage >= 90% | Countries with DTP3 Coverage < 90% | All countries |
|---|---|---|---|
| Number of countries that received performance payments* | 11 | 5 | 16 |
| PBF rewards | US$30,441,080 (87.5%) | US$4,331,730 (12.5%) | US$34,772,810 |
| Disbursements received | US$12,498,280 (80%) | US$3,130,802 (20%) | US$15,629,082 |
| Average size of PBF payment | US$1,217,643 (IQR: US$940,000 – US$4,526,000) | US$618,819 (IQR: US$73,020 – US$2,289,480) | US$1,086,650 (IQR: US$730,000 – US$4,150,000) |

*IQR: inter-quartile range*

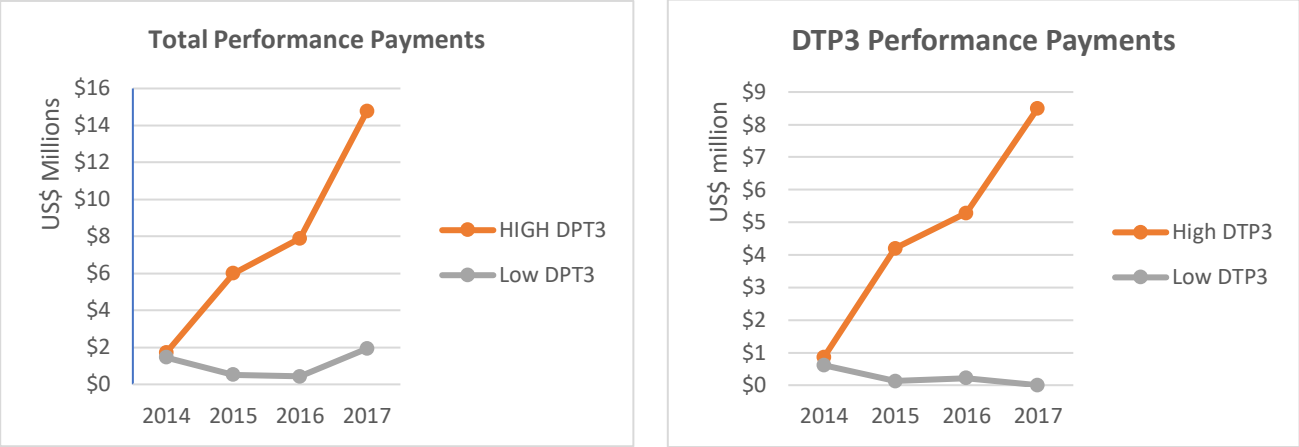
There was also a difference in payment trends across country groups and type of payments. For total performance payments, the high coverage group showed a rising trend and collectively received more than the low coverage group in every year between 2014 and 2017 while the low coverage group showed a declining trend in total performance payments between 2014 and 2016 but showed an increase between 2016 and 2017 (Figure 6, left graphic).

For DTP3 performance payments, the high-country groups also showed a rising trend in terms of earning performance payments. However, the 11 high coverage countries received a total of 24

rewards for DTP3 but only 7 of these payments were for increasing coverage (the remaining 17 kept coverage at least at 90% coverage levels).

Unlike for total performance payments, DTP3 payments to low coverage countries continued to decline throughout the period with **no country from the low coverage group receiving a performance payment for increased DTP3 coverage in 2017** (Figure 6, right graphic). There was an increase in 2017 in performance payments for MCV1 coverage in the low coverage group (Annex 6).

**Figure 6. Total performance payments and DTP3 performance payments**




**Likelihood of eligibility for performance payments**

**Key findings: Our analysis found that countries with high baseline coverage are 4.5 times more likely to be eligible for *any* performance payment than countries with baseline coverage below 90%. This finding shows that the PBF system is not balanced and disadvantages countries with lower baseline coverage levels.**

We estimated the likelihood of both high coverage countries and low coverage countries being eligible for performance payments. We then used these estimates to calculate the relative likelihood of eligibility for high coverage countries compared with low coverage countries. Our analysis shows that for the overall period 2014-2017, high coverage countries were:

- 6.9 times more likely to be eligible for DTP3 performance payments (Table 9)
- 3.21 times more likely to be eligible for the geographic equity performance payment than low coverage countries are for the MCV payment (Table 10)
- 4.5 times more likely to be eligible for *any* performance payment (i.e. DTP3 and/or MCV or geographic equity performance payments) (Table 11).

Thus, high coverage countries have a much higher likelihood of qualifying for PBF rewards (see also Figure 7).[79]

[79] Preliminary 2018 performance data was shared with us at the time of writing the report. It shows the same pattern. High DPT3 coverage countries had a success rate of 80% and an average award amount of US$ 1 million, while low DTP3 coverage countries had a success rate of 5% and an average award amount of US$36,400.

**Table 9: Likelihood of being eligible for DTP3 performance payments**

| DTP3 Performance Payments | | | | | |
|---|---|---|---|---|---|
| **High DTP3 Coverage Countries** | **2014** | **2015** | **2016** | **2017** | **Total** |
| Number of countries assessed | 1 | 7 | 9 | 11 | 28 |
| Number eligible | 1 | 6 | 7 | 9 | 23 |
| Likelihood of being eligible for performance payment | 1.00 | 0.86 | 0.78 | 0.82 | 0.82 |
| **Low DTP3 Countries** | **2014** | **2015** | **2016** | **2017** | **Total** |
| Number of countries assessed | 4 | 9 | 12 | 17 | 42 |
| Number eligible | 1 | 1 | 2 | 1 | 5 |
| Likelihood of being eligible for performance payment | 0.25 | 0.11 | 0.17 | 0.06 | 0.12 |
| **Relative likelihood of being eligible (High/Low)** | **4.00** | **7.71** | **4.67** | **13.91** | **6.90** |

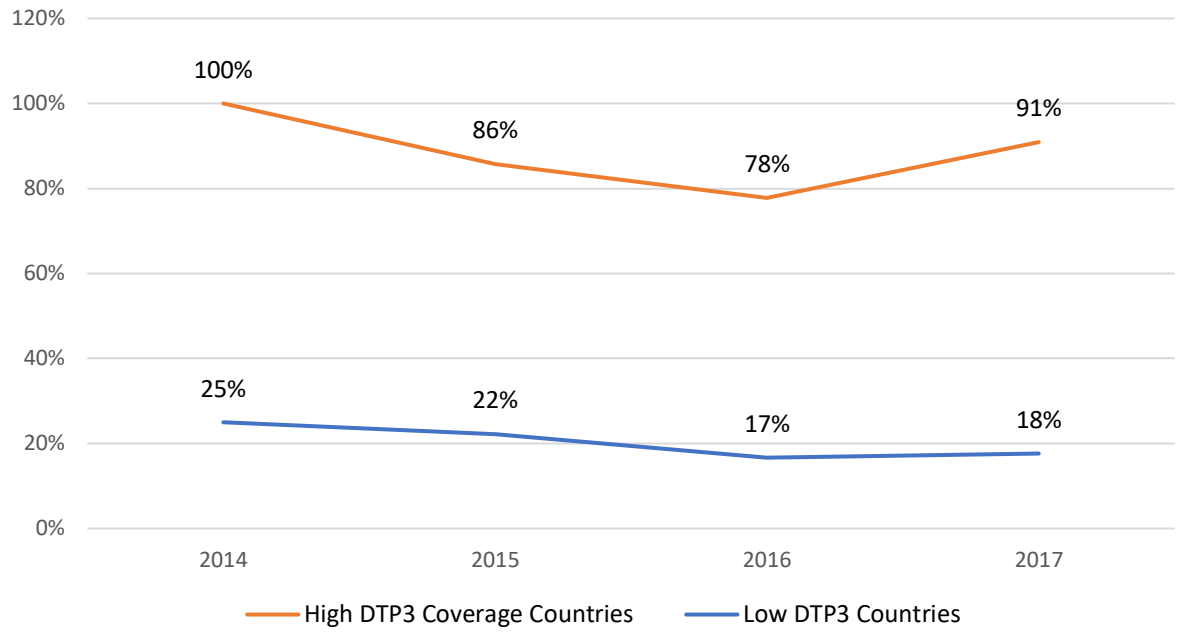
**Table 10: Likelihood of being eligible for MCV or geographic equity performance payments**

| MCV/Geo Performance Payments | | | | | |
|---|---|---|---|---|---|
| **High DTP3 Coverage Countries** | **2014** | **2015** | **2016** | **2017** | **Total** |
| Number of countries assessed | 1 | 7 | 9 | 11 | 28 |
| Number eligible | 1 | 4 | 3 | 7 | 15 |
| Likelihood of being eligible for performance payment | 1.00 | 0.57 | 0.33 | 0.64 | 0.54 |
| **Low DTP3 Countries** | **2014** | **2015** | **2016** | **2017** | **Total** |
| Number of countries assessed | 4 | 9 | 12 | 17 | 42 |
| Number eligible | 1 | 2 | 1 | 3 | 7 |
| Likelihood of being eligible for performance payment | 0.25 | 0.22 | 0.08 | 0.18 | 0.17 |
| **Relative likelihood of being eligible (High/Low)** | **4.00** | **2.57** | **4.00** | **3.61** | **3.21** |

**Table 11: Comparison of the likelihood of being eligible for any (DTP3 and/or MCV/Geo) performance payment by level of DTP3 coverage**

| Any Performance Payments (DTP3/MCV/Geo) | | | | | |
|---|---|---|---|---|---|
| **High DTP3 Coverage Countries** | 2014 | 2015 | 2016 | 2017 | Total |
| Number of countries assessed | 1 | 7 | 9 | 11 | 28 |
| Number eligible | 1 | 6 | 7 | 10 | 24 |
| Likelihood of being eligible for performance payments | 1.00 | 0.86 | 0.78 | 0.91 | 0.86 |
| **Low DTP3 Countries** | 2014 | 2015 | 2016 | 2017 | Total |
| Number of countries assessed | 4 | 9 | 12 | 17 | 42 |
| Number eligible | 1 | 2 | 2 | 3 | 8 |
| Likelihood of being eligible for performance payments | 0.25 | 0.22 | 0.17 | 0.18 | 0.19 |
| **Relative likelihood of being eligible (High/Low)** | 4.00 | 3.86 | 4.67 | 5.15 | 4.50 |

**Figure 7. Likelihood of qualifying for any type of performance payment**



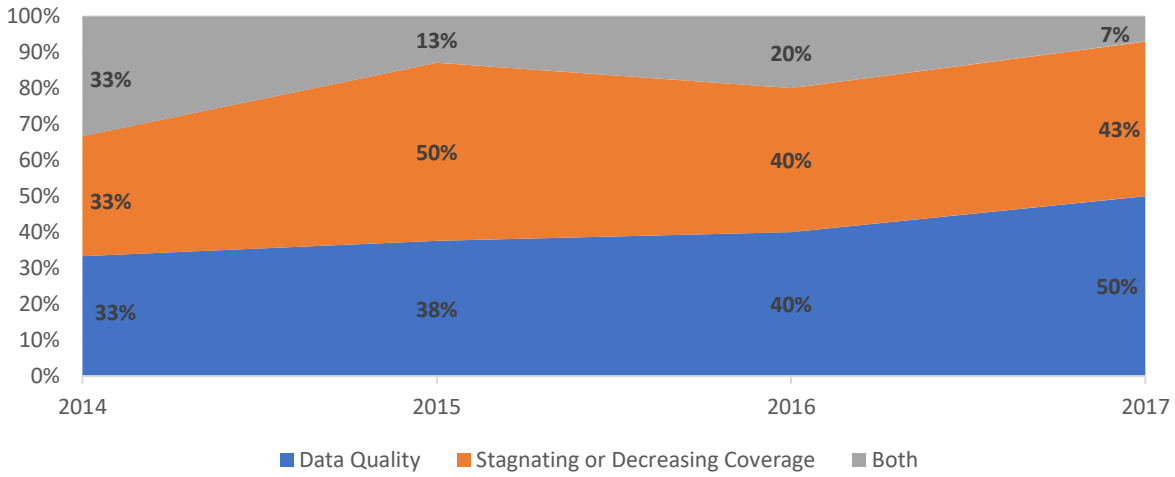

**Reasons for ineligibility**

**Key findings: In addition to lack of progress in terms of immunization coverage, the verification mechanism has prevented countries with weak data from accessing performance rewards.**

Gavi's published data show the reasons why countries have been found ineligible for performance payments (Figure 8). In 2017, 50% of cases of ineligibility were due to poor data quality, 43% due to stagnating/decreasing coverage, and 7% to both.

Thus, the verification mechanism has prevented countries with weak data from accessing performance rewards. As we highlighted in Section 4.2, many countries do not qualify for the performance payment due to the design of the measurement and verification system, a major weakness of the PBF's overall design. Since 2015, countries have been able to choose WUENIC data for their baselines; as such, the challenge linked to the use of country data might to some extent have been mitigated. However, many countries may still stick to their own administrative data as data are often highly politicized.

**Figure 8. Reasons why countries are ineligible for performance payments, 2014-2017**




**Role of geography and fragility in ineligibility**

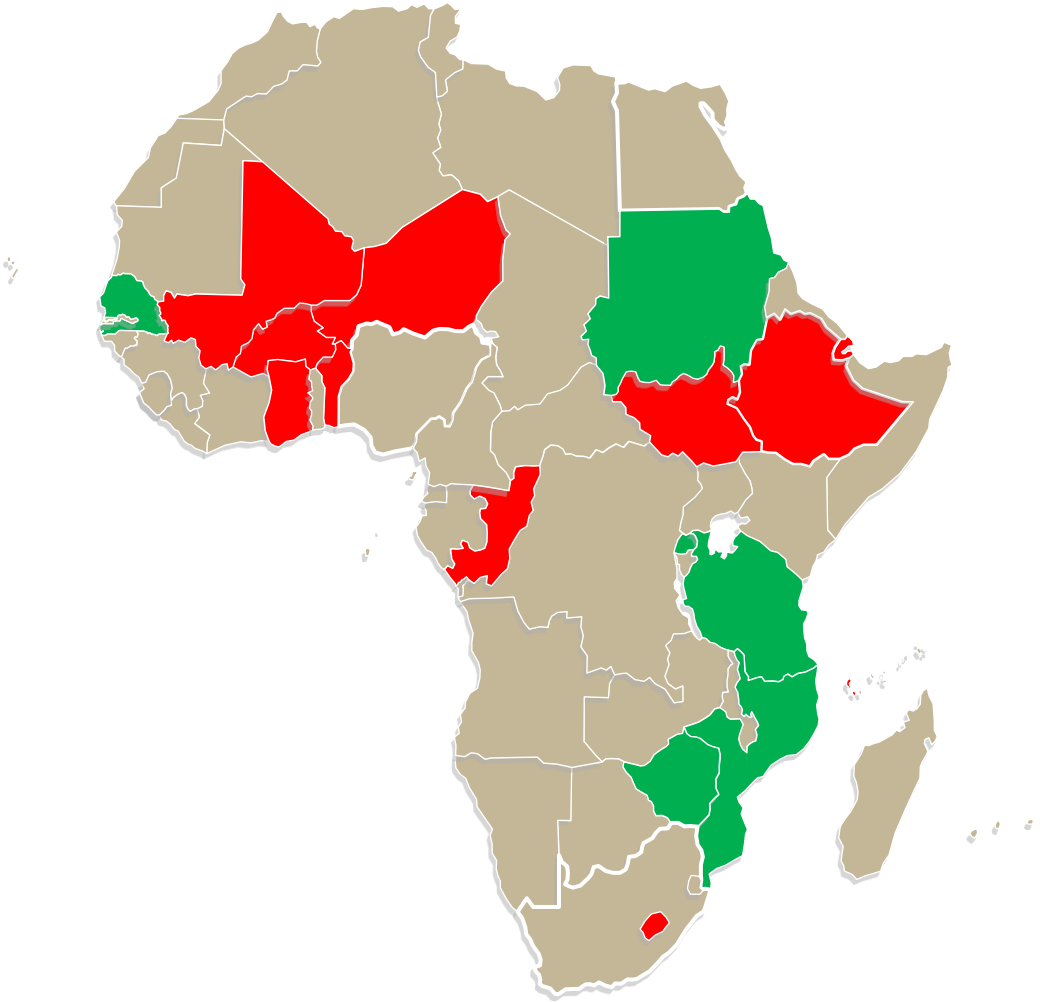
**Key findings: Only two French-speaking countries from Africa ever qualified for the performance payment, and only 4 out of 18 countries from Gavi's fragility list ever did.**

In 2017, 17 out of 28 countries assessed for eligibility (60%) were in sub-Saharan Africa. Figure 9 shows which African countries were assessed for eligibility and which received the performance payment in 2017. Over half of all assessed countries were ineligible, mostly (70% of the time) because of poor data quality. The only French-speaking countries from sub-Saharan Africa that ever received the performance payment are Burundi and Senegal, with Burundi being the only country that received the payment multiple times.

In terms of fragility, only four out of the 18 countries on Gavi's fragility list have ever qualified for a performance reward: Burundi, Sudan, Solomon Islands, and Zimbabwe.[80] Afghanistan did not earn rewards but received funds to improve the data system under the country tailored approach.

[80] At the time of writing this report, 18 countries were on Gavi's fragility list; since then, Nigeria and Ethiopia have been removed.

**Figure 9. Countries assessed for eligibility for performance payments in 2017.** Countries in green received the payment, those in red did not.



## 6.2 Did the PBF approach incentivize more action towards improved immunization coverage and equity at country level? If so, what kind of action?

**Key findings: Our study found mixed results on whether the PBF had a motivational effect. Countries that benefitted from the PBF in the past reported that the PBF incentivized better planning and implementation and helped them to focus on the equity agenda. However, other countries reported that the PBF did not motivate them.**

A range of country stakeholders as well as members of Gavi's Country Programmes team argued that PBF has helped countries to start thinking about prioritization, especially in terms of equity. It also helped them to try harder to improve coverage in regions/districts with lower coverage. Countries like Nicaragua used the performance reward to improve equity and strategic planning. The core funding from the HSS grant was programmed so that low-coverage districts were further prioritized. This prioritization also occurred in other countries like Burundi, Lao PDR, Solomon Islands, Tanzania, and Zimbabwe. In Burundi, Gavi's performance payment was used to pilot a results-based financing mechanism at community level. This pilot was successful and led to the development of a community-based strategy. There is now also a three-year scale-up process to roll it out in all 18 provinces.

One country representative from a country in Sub-Sahara Africa (SSA), for example, reported:

> "PBF is really motivational. We used the funds for outreach work in low-coverage districts, and we also used it to fund additional support in districts where we see stagnating or even declining coverage- these are not necessarily the ones that are a priority of Gavi's general HSS grant. In this way, the PBF was really complementary."

Another country representative from SSA was even more concrete:

> "The PBF encourages [us] to improve our service delivery – cold chain, distribution of vaccines and other things. It also helps in reaching remote areas**.** We also improved the reporting at district level to qualify for PBF."

An EPI manager from Asia reported:

> "PBF is quite a clever mechanism to ensure that money is spent where it is needed most. Some districts need better storage, motorbikes, training of disseminators. We focus on districts with weak coverage. We improved supervision, data quality, and train local people to deliver the communication to increase demand for vaccinate. Coverage increased in many districts, which benefitted from the extra PBF reward."

Other country stakeholders highlighted that Gavi's PBF helped to introduce a performance-based service delivery culture, which shifted the focus from inputs towards measuring outputs. In one country, this also helped other funders with their performance-based programs.

Stakeholders also reported that the incentive trickles down to lower levels of the country and that it has changed the management approach:

> "We have quarterly EPI meetings at which all districts are represented. We together look at the data and discuss how we can improve coverage in districts that are lacking behind, also to ensure that we qualify for Gavi's performance payment in the next year. It really brings together the EPI team beyond the national level. In addition, the way we manage the EPI has changed. We use much more of performance-based management approach in the country."

Another SSA EPI manager explained that the RED (reach every district) by WHO was modified and tailored to the local country context: "We revised a lot of guidelines and the RED plan to win the performance payment. This is really important for a country like us."

Some SCMs raised doubts during interviews about the effect of the incentive at subnational levels. However, a country representative from Zimbabwe and a few other countries reported during the interviews that the PBF helped to improve the performance at subnational level and that the effect trickled down.

It is questionable whether countries with high baseline coverage, which are much more likely than low baseline coverage countries to receive the rewards, always try harder to get the reward (it is already rewarded for having high coverage at baseline).

> "[The] country didn't have to do anything different to get the reward, it didn't motivate them to do anything differently at all, there is a lot of unspent HSS money already, to be honest."

> "The PBF support comes – you do not need to plan. It is very helpful and not so stressful to implement as the more general HSS because is very important to the country."

With respect to French-speaking African countries, Gavi Secretariat members also argued that there might be less motivation: "[I am] not sure that PBF is very useful in Francophone countries - it is not a motivation for them to develop a better plan."

As we highlighted, countries may have limited knowledge of Gavi's PBF but key informants from the country level gave additional explanations of why the performance payment did not incentivize them. Some countries indicated that the burn rate of the overall HSS grant is quite low—therefore it is not a high priority to win and successfully implement the PBF grant. Others said that the size of the payment was not large enough to motivate them. Yet another country reported that the funding was not sufficiently aligned and harmonized ("mainstreamed") with other results-based initiatives. The following quotations from country representatives illustrates the broad range of reasons why not all countries were incentivized by the PBF:

> "Our goal was to use the HSS grant well. Qualifying for PBF was not one of our objectives".

> "PBF was not mainstreamed and political interest was low. HMIS reporting systems were weak, there were no good mechanisms to verify data, and there was lack of trust about the reliability of the data."

> "We know that we can access this extra money, but it is not so much. It might have marginal impact. It would not have an impact on coverage and equity. We never did anything special to get it. It is very helpful to fill gaps in our budget and it is not stressful to implement."

> "First, data quality is poor here. It is hard to measure performance. Second, country capacity to implement and use funds is limited. We face a real administrative and financial limitation in processes making it hard to use funds. PBF is welcome but we struggle to even implement the overall Gavi HSS grant. So PBF really is not a big incentive".

Overall, the picture on the motivational aspect is mixed. In a range of countries, particularly in those that qualified for performance payments, the PBF incentivized better planning and implementation, and it appears that it was especially helpful to boost the equity agenda. A range of countries as a result of the PBF focused more on coverage in districts and provinces that were lagging behind. At the same time, there are countries that reported that the PBF did not motivate them. There are several reasons (mentioned above), such as the size of the payment, low HSS grant burn rates, or the lack of alignment, which have to do with the overall design of Gavi's PBF model.

## 6.3 To what extent did Gavi's PBF contribute to improvements in EPI coverage at country level? What were the major factors influencing the achievement of these results?

**Several countries that received a PBF payment showed rising immunization coverage and improvements in equity but a causal link between the payment and changes in coverage/equity cannot be proved.**

It is methodologically very complex to attribute changes in coverage or equity specifically to the PBF, and causal attribution is beyond the scope of this review. Given that there has been no formal impact evaluation using randomization and controls, it is not possible to specifically attribute changes in coverage/equity specifically to the PBF itself. However, our country deep dives provide evidence to suggest that Gavi's PBF to some extent contributed to improved coverage and equity. For Burundi, Lao, and Tanzania we found some evidence that the PBF has likely contributed to improvements. There is little evidence for Sudan, and Ethiopia never earned a PBF payment.

*Lao PDR*

The interviewed country-level stakeholders in Lao argued that the PBF helped to increase coverage and equity:

> "PBF has absolutely contributed to equity and coverage. It allows us to access funds. Based on this, we can act quickly to increase coverage in ethnic groups, rural areas; [and] children in the bush and slums."

A review of immunization coverage indicators shows that Lao increased DTP3 coverage from 78% in 2011 to 89% in 2015, and MCV 1 coverage increased from 69% to 88% in 2015. However, in 2016, both indicators dropped substantially. One reason for this drop in coverage was that a polio outbreak diverted resources from the routine immunization. In addition, some of the campaigns were run in an integrated manner (e.g. polio and measles), which led to challenges in terms of data collection and lower registration number for measles and DTP.

The interviewed country stakeholders reported that PBF funding contributed to increased coverage in certain districts. While the performance of districts significantly improved since 2011, there was a

substantial drop in 2016 in line with drop in coverage at national level (this indicator rose slightly to 64% in 2017).

*Tanzania*

The first PBF disbursement to Tanzania was only made in December 2017, so we do not expect that an impact from this performance payment can be seen yet. However, key informants reported that the focus on equity stimulated discussions at the subnational level on ways to ensure that no child was left out. A review of immunization coverage indicators shows that Tanzania increased DTP3 coverage from 91% in 2013 to 97% in 2016.[81] According to WUENIC data, the percentage of districts reporting DTP3 coverage greater than 80% increased from 79% in 2013 to 92% in 2016. MCV1 coverage stayed the same at 99% between 2013 and 2015 but dropped to 90% in 2016. This drop in coverage has been attributed to procurement delays that happened at the country level.

According to a recent data review conducted by Gavi (April 2018), there remains substantial variation in immunization performance across districts. While the number of districts with coverage below 80% increased slightly in 2017, many more districts achieved coverage levels of more than 90%. However, in 2017, nearly half (45%) of the districts had a DTP3 coverage of over 100%, which shows that there are still significant measurement issues.

*Burundi*

A review of immunization coverage indicators show that Burundi maintained its DTP3 coverage above 90% between 2015 and 2017. This is remarkable because Burundi suffered heavily from the 2015 political crisis in 2015, with many donors withdrawing or freezing their support. Because Gavi managed to provide its cash-based support through UNICEF (following negotiations with the government over almost one year), it allowed the country to maintain its coverage high despite the risk of collapsing entirely.[82] As such, Burundi represents an important example of how performance payments can effectively be used in fragile contexts.[83]

A review of immunization coverage indicators show that Burundi maintained its DTP3 coverage above 90% between 2012 and 2016 (Table 3). However, equity (defined as the percentage of districts with at least 80% DTP3 coverage) fell from 93% in 2012 and 2013 to 84% in 2014 and 89% in 2015. However, it returned again to 93% in 2016.

[81] Interviewees did mention three changes that had happened, but these were in the context of the overall HSS grant and EPI program, and it is hard to know whether or how much the PBF made a specific contribution. These changes mentioned by key informants were: (1) development of a country vision of what to do with the funds, (2) improvements to the vaccine storage capacity from being able to store supplies for one month to now three months, and (3) adaptation of the WHO's RED (Reaching Every District) strategy guidelines (and other guidelines) to be more applicable to Tanzania.
[82] Also, the (second) performance payment was fully integrated with the third core tranche of the HSS 2 grant.
[83] Through its general HSS grant, Gavi also supported a country-wide performance-based financing scheme, which, according to Burundi's latest available country joint appraisal (2016), helped with increasing the use of immunization services. Specifically, the joint appraisal states: "PBF is noted as one of the important factors in increasing the use of health services, including vaccination. It is important to make an impact study of the clinical effects of PBF on indicators of vaccination." See: https://www.gavi.org/country/burundi/documents/

### 6.4. What have been the unanticipated consequences of PBF at country and global levels (if any)?

**Key findings: We found little evidence on major unintended consequences. In one country concerns were raised about over-reporting as a negative unintended effect. A positive unintended effect of the PBF might be that the need for country data systems was further discussed.**

The available documentation and the key informant interviews did not point to any major unanticipated consequences of the PBF model. The challenge related to the measurement of performance, however, may have further reinforced the need to improve country data systems. This is a positive unanticipated effect as it help to channel funding to countries like Afghanistan under the country tailored approach (and now to Ethiopia, under the fragility policy).

In one deep dive country (Lao PDR), one stakeholder raised concerns about the quality of data and the potential to overreport: "There needs to be a quality check. The data needs to be validated and monitored. We need more supervision. We need better impact evaluation. Drawbacks are in quality of data and close monitoring." Stakeholders thus demanded a more independent assessment of the impacts of PBF through Gavi.

## 7. Lessons learned

**Key findings: Gavi should keep its PBF model if it ensures learning and is willing to take a risk. If not, the PBF model should be ended.**

Our review suggests that Gavi should continue with a PBF model. However, **Gavi should only keep the PBF model under two conditions**. The first is that it **ensures learning and provides the conditions for learning** within the Secretariat and the whole alliance. The Secretariat needs to develop the necessary tools to effectively monitor the implementation and results of the PBF. Based on the improved monitoring, it should regularly report to the Board and the PPC to allow for discussion and learning. Second, Gavi should also decide **if it is willing to take risk**. In 2011, the risks associated with PBF were only discussed briefly in the respective PPC and Board documents. We believe that Gavi's continued testing of the PBF model will require the organization to be comfortable in accepting a certain level of risk. All PBF models run the risk of creating perverse incentives—in particular, the risk of over-reporting (inflating results in order to receive reward payments). To remove this risk almost entirely requires costly, highly intensive, external verification systems that can end up constituting a huge proportion of the total costs of the PBF scheme. We believe that Gavi's approach of investing in national data systems, rather than external verification, is much more valuable over the long run. If Gavi cannot put in place the conditions for learning, and if it is not willing to take a risk, the PBF model should be ended.

**We believe that Gavi should keep the model for three reasons:** First, Gavi has been at the forefront of PBF and entities such as the GFF will continue to implement it. The interest in PBF among donors and countries is growing, and we are still in a highly active learning phase. It is critical for Gavi to be in the PBF "learning arena." As a group of Rwandan health experts noted (Paulin Basinga and colleagues), PBF has great potential, and while the evidence for its effectiveness overall has been

mixed, we should continue to learn what works best rather abandoning the approach.[84] They argue that: "Continuous checking and integration of the PBF approach is needed during implementation and this should be informed by operational research aimed at aligning PBF with the existing health system." Overall, we believe that Gavi's approach of investing in national data systems is thus highly valuable over the long run, including for the PBF.[85]

Second, we found that many countries are motivated by the PBF model to increase their immunization efforts (Section 6).

Third, the PBF model is also valued by countries that have received the payment; countries that we have interviewed gave the model high ratings for relevance and overall success (Table 12). The average score across nine countries (from 11 interviews) was 3.5 for relevance (using a scoring scale of 1-4, where 4 is highest relevance) and 7 for success (using a scoring scale of 1-10, where 10 is highest relevance).

**Table 12. Ratings by key informants of the relevance and overall success of Gavi's PBF model**

| Country | No. interviews | Average score for relevance of PBF (scale of 1-4) | Average score for overall success of PBF (scale of 1-10) |
|---|---|---|---|
| Burundi | 3 | 3,5 | 9 |
| DPRK | 1 | 4 | 6,5 |
| Honduras | 1 | 4 | 9 |
| Laos | 4 | 3,5 | 7,1 |
| Mozambique | 1 | 2 | 5 |
| Solomon Islands | 2 | 3 | 8,5 |
| Sudan | 1 | 4 | 8 |
| Tanzania | 3 | 4 | 8,7 |
| Timor Leste | 2 | 3 | 1,5 |
| Zimbabwe | 3 | 4 | 8 |
| **All countries** | 21 | **3,5** | **7,13** |

[84] Basinga P, et al. Performance-based financing: the need for more research. *Bull World Health Organ* 2011 Sep 1; 89(9): 698–699.

[85] Our analysis did not find evidence that countries that had received **Health Information System** (HIS) support from Gavi were more likely to receive a PBF reward payment (Annex 8). However, this analysis faced limitations in terms of data - HIS disbursement data was unavailable and thus budget data was used which is a serious constraint. While we thus cannot prove that Gavi's HIS support *led* to higher success, we think that these payments will pay off and contribute to better performance measurement.

However, Gavi's PBF is not a panacea, and it is has not worked in all countries. The PBF approach needs significant revisions, both in terms of design and process. One major process weakness has been the poor communication and guidance about the model. It will also be critical to ensure a meaningful participation of countries in the development of any new PBF model. In the implementation of such a revised model, clear communication to all countries will be a necessary precondition for successful implementation.

During the redesign process, the current model should continue in its current form. Abruptly stopping the model would likely foster confusion, disruption, and unpredictability at country level.

**Key findings: The PBF design needs to be changed. It suffers from being too ambitious and its underlying assumptions are flawed.**

The outcome focus is too difficult for many countries, and more time is needed to achieve the outcomes (the initial evaluation of the SMI also found that the short timeframe was challenging).

There is a long delay in receiving the payment, and the model does also not take into account the overall HSS grant performance. Low coverage country groups include very different countries; some low coverage countries received support based on the fragility policy (e.g. Afghanistan) but others did not. This is not a good overall strategy (it is inconsistent and relies on ad hoc decision making).

In addition, some countries are not motivated to qualify for performance payments, which to some extent is also due to communication challenges. In addition, there is evidence that the payment is too small. The ISS review already suggested that the incentive payment is too small, and this design weakness was also reflected in the PPC PBF paper from September 2010.

**Key findings: The PBF design is too sensitive to measurement errors.**

The verification is not working well; there are major issues with the data. We ran three different hypothetical models to see what could help to alleviate these data and verification issues.

First, we assessed whether a "higher tolerance level" (allowing a 10% rather than 5% difference between WUENIC data and administrative data) would have made a difference in the 2014-17 timeframe. However, it would not have made much difference. Most countries that could not access the PBF because of data quality issues have huge differences between WUENIC data and administrative data (e.g. 23% for Benin, 26% for Mali, 25% for Niger).

Second, we compared each country's actual payment experience with the payment experience under a hypothetical WUENIC option.[86] The results show that under the WUENIC option, 17 countries (representing 31 PBF awards) would have had a different payment experience (Table 13). Fifteen of the 17 countries would have received at least one (range: 1 to 4) additional PBF award, one country (Burundi) would have lost one PBF award, and one country (Zimbabwe) would have lost one DTP3 award but gained one equity award. Both countries that would have lost a PBF award have high baseline DTP3 coverage and would have lost the equity award. Of the 15 countries that would have received additional PBF awards, 13 (representing 26 awards) had low DTP3 coverage at baseline while two countries (representing two awards) had high DTP3 coverage at baseline.

However, while the new WUENIC model may thus help to some extent, more reforms will be needed. Using WUENIC data for baseline and progress assessments will not be a magic bullet.

**Table 13. Comparison of payments under WUENIC option with countries actual payment experience.**

| Performance under WUENIC compared to actual | Countries with DTP3 Coverage >= 90% | Countries with DTP3 Coverage < 90% | All countries |
|---|---|---|---|
| Would qualify for at least one additional PBF but NOT lose any (2014-2017) | 2 | 13 | 15 |
| Would lose a PBF award they received but NOT gain an additional award | 1 | - | 1 |
| Would gain an award in one year and lose an award in another | 1 | - | 1 |
| No Change | 8 | 6 | 14 |

| Performance under WUENIC compared to actual | Countries with DTP3 Coverage >= 90% | Countries with DTP3 Coverage < 90% |
|---|---|---|
| Would qualify for at least one additional PBF but NOT lose any (2014-2017) | Lesotho, Sudan | Afghanistan, Benin, Burkina Faso, Cambodia, Comoros, Ethiopia, Haiti, Liberia, Mali, Niger, Senegal, Timor-Leste, Yemen |
| Would lose a PBF award they received but NOT gain an additional award | Burundi | - |
| Would gain an award in one year and lose an award in another | Zimbabwe | - |

[86] Under this option, countries are assessed based on WUENIC data only. However, countries were only assessed for eligibility in the years for which they were actually assessed for eligibility based on data received from Gavi.

| | | |
|---|---|---|
| No Change | Bangladesh, DPRK, Ghana, Honduras, Nicaragua, Rwanda, Solomon Islands, United Republic of Tanzania | Congo, Djibouti, Lao People's Democratic Republic, Mozambique, Papua New Guinea, South Sudan |

Third, we modeled a scenario where countries with low baseline DTP3 coverage qualified for an equity bonus if they showed an increase in the percentage of districts with DTP3 coverage ≥80%, regardless of their performance on other metrics (we note that this is an alternative equity measure than the one currently used by Gavi for high-coverage countries). In this scenario, 12 out of the 19 low DTP3 countries would have qualified for at least one equity PBF bonus while 7 countries would not have qualified. Of these, one country (Ethiopia) would have qualified for three equity bonuses, three countries (Afghanistan, Benin, and Lao PDR) would have qualified for two equity bonuses, and eight countries (Cambodia, Comoros, Haiti, Liberia, Mali, Niger, Senegal, and Timor-Leste) would have qualified for one equity bonus. Burkina Faso, Congo, Djibouti, Mozambique, Papua New Guinea, South Sudan, and Yemen would not have qualified for any equity bonuses under this scenario. Thus, several countries that are currently not being rewarded for equity performance would have qualified for such an award. This shows the potential to introduce equity indicators to the lower-coverage group.

**Key findings: There have been problems with communication about the PBF model. There have also been challenges in aligning Gavi's PBF model with other PBF mechanisms in-country. There has been little in the way of learning.**

Overall, our review found that there is generally very little knowledge at country level about the PBF model. Countries often do not know the model even exists, let alone how to qualify for the reward payment. PBF has not been well communicated to country partners. We also found little in the way of alignment between Gavi's PBF and other PBF/RBF schemes in country, a situation that could be hindering overall effectiveness and impact. Lastly, given that PBF as a model remains "experimental" (our review of the literature showed that its effectiveness remains unclear), its use should always be accompanied by evaluation and learning—yet we found that Gavi's PBF was not embedded in a learning environment.

## 8. Recommendations

In this final section, we address the question "if a PBF mechanism is still required, how should Gavi restructure it to make it more relevant, efficient, and effective to contribute to achieving Gavi's 2016-20 strategy?" As mentioned above, we believe that if a learning environment is established and Gavi is willing to take a risk, the PBF model should be continued but with important changes to its design.

***Design recommendations***

**Recommendation 1: The PBF model needs to be redesigned, since it does not work for all countries. We recommend that the model should have three country groups: a high coverage group (DTP3 coverage of at least 90% at baseline), a medium coverage group (70-89% DTP3 coverage at baseline), and a low coverage group (under 70% at baseline). The current PBF model works well enough for the high coverage group and can be kept. For the medium coverage (70-89%) group, we propose new indicators and decision rules to qualify for performance payments. For the low coverage group (below 70%), PBF should not be used—instead, countries need to first build their health systems capacity, including health information systems.**

▪ *For the high coverage group (≥90% DTP3 coverage at baseline)*: The current PBF model rewards countries with high coverage to maintain coverage levels; the existing model is good enough for these countries and we believe it can be kept.

▪ *For the medium coverage group (70-89% DTP3 coverage at baseline)*: For this group, we propose a new approach involving three new indicators: (a) a standardized systems indicator that cuts across all countries from this group (measuring stock-outs), (b) a second, country-specific immunization process/systems indicator chosen from a country's grant performance indicators, which reflect Gavi's investments in HSS in individual countries, and (c) an equity indicator. Country performance for these three indicators should not be measured as "pass or fail"—instead progress in these indicators should be measured continuously:

- **Standardized systems indicator:** Intermediate system and/or process indicators rather than coverage indicators should be used to incentivize and reward the medium coverage group. There is strong evidence showing that selected systems/process indicators contribute to coverage and equity. One specific system indicator that could be used for this purpose is *vaccine stock-outs*, as measured, for example, by the proportion of facilities with full availability of all or a selected set of tracer vaccines and immunization supplies over a resupply period. Gavi routinely receives this information as this is a core intermediate indicator from its grant performance framework.[87] Monitoring of stock-outs is routinely conducted by countries as part of the Joint Reporting Process. However, the country-

[87] Gavi, the Vaccine Alliance: Considerations for countries on targeting Gavi investments to achieve immunisation outcomes. Focus Area Immunisation Supply chain. Revised Version May 2018. Available at: https://www.gavi.org/support/process/apply/hss/

reported data would need to be verified, which will involve additional transaction costs (see Recommendation 3 below).[88]

- **Country-specific system/process indicator**: There should be a second systems or process indicator that would strongly reflect Gavi's HSS investments in individual countries. This second indicator would also be part of the grant performance framework. It could also be a core indicator (like the indicator on stockouts), or a tailored indicator. This country-specific indicator would be selected *by countries themselves* in dialogue with the Gavi secretariat (examples include the proportion of planned immunization outreach sessions conducted; the proportion of health facilities with at least one qualified/trained vaccine provider). Under this second indicator, Gavi could also reward improvements in data accuracy.
- **Equity indicator:** For the medium coverage group, performance payments should also be made conditional on improvements in equity. Gavi could use the equity indicators from the high coverage group, which is in line with the Global Vaccine Action Plan, though we suggest a more radical approach (see Recommendation 2 below).[89] Rather than using this indicator in a "pass or fail" way, continuous progress should be measured.

▪ *For countries with <70% DTP3 coverage at baseline:* PBF should not be used. Instead, Gavi should support these countries to develop their systems first (infrastructure, data systems, etc.) before being included in the PBF scheme. We arrive at this recommendation based on thorough triangulation of our results. First, the low coverage group comprises 11 countries from the list of countries that are eligible to apply for Gavi support in 2018.[90] Nine of these 11 countries are on Gavi's 2018 list of fragile states, so rather than being low-resource settings only, these countries face an additional set of major systemic challenges. The conditions in these countries are extremely difficult, with large-scale conflict, high risk of disease outbreaks, and very weak or non-committed state actors. Findings from our literature review show that PBF is difficult to implement in fragile states, with a range of studies concluding that country fragility is a major barrier to the success of PBF. The barrier is due to problems such as weak health information systems, governance, and leadership. Second, Gavi's own PBF data shows that the implementation of PBF models is difficult for the low coverage group. No country with less than 70% DTP3 coverage at baseline qualified for a performance payment between 2014 and 2017. Only four fragile countries qualified for a PBF performance payment in this period. Gavi also already recognized the challenges associated with these fragile settings, and provided the performance payments to fragile countries like Afghanistan and Somalia under the country tailored approach, although they did not qualify for these payments. Third, key experts that we interviewed highlighted the complexities in fragile countries, and argued that the implementation of PBF models in such countries is often overly ambitious and inappropriate to multiple challenges. We thus believe that Gavi should not use PBF for the low coverage group. Once the countries are ready, and have been accredited, they can be enrolled into the PBF scheme.

[88] Going forward, it would be critical for Gavi to further standardize the way that this stock-out indicator is monitored across countries.
[89] http://www.who.int/immunization/global_vaccine_action_plan/GVAP_doc_2011_2020/en/
[90] https://www.gavi.org/support/sustainability/countries-eligible-for-support/

The World Bank uses such an accreditation model in its results-based programs (which are focused on health facilities and as such have a different focus than Gavi's PBF model).

We acknowledge that the recommended segmentation involves different mechanisms for each group by which the incentive could link to performance. High coverage countries would continue to be rewarded for maintaining coverage at a minimum of 90%, while countries from the medium coverage group need to show improvements on key system indicators. However, maintaining coverage at a high level (at least 90%) signals strong continuous performance across the system and countries should be rewarded for this achievement. For the medium coverage group, a focus on system indicators is much more appropriate and – together with an improved verification approach – should lead to more balanced PBF results, i.e. to a model that works for all countries even if different mechanisms are at play.[91] Overall, we think acknowledging the different country conditions is more important than designing one model that cuts across all countries.

**Recommendation 2:** We believe Gavi should consider a more radical measure for equity for both the high coverage and the medium coverage groups. Gavi's goal of increasing *national* coverage is not necessarily in line with its current equity efforts as geographic areas that already lag behind might receive even less attention as they are the hardest to reach.[92] We suggest that the PBF model should be targeted at increasing coverage in three areas of very low coverage ("hot spots"). Periodic population-based surveys provide data that could pinpoint parts of the country that are under-served hot spots. Gavi could incentivize countries to increase their attention to these areas. To measure performance, Gavi would require additional monitoring of performance in these hot spots as a supplement to more standard, nationwide measures of the immunization system and coverage. Gavi could require countries to focus on a subset of those regions or districts with the largest percentage or number of zero-dose (zero-antigen) children who never received a single dose of a vaccine. The performance payment would then be paid based on the additional number of previously zero-dosed children.[93] As an alternative to rewarding additional zero-dosed children immunized in these hot spots, Gavi could reward the additional number of children immunized.

***Implementation recommendations***

**Recommendation 3: On data verification, the current system is working well enough for the high-performing countries, but a new system is needed for the 70-89% coverage group.** For the 70-89% coverage group, a new mechanism will need to be put in place to verify the reported system indicators, including stock-outs. Gavi should avoid using very costly and labor-intensive verification approaches (e.g., those used by the World Bank or the SMI).

[91] For both the high and medium coverage groups, we suggest that there are no changes in the relationship between the programmable payment and the performance payment - the current model ensures sufficient predictability.
[92] In addition to the spatial dimension of equity, there are other dimensions. Groups from the lowest socioeconomic quintile, and/or marginalized and discriminated groups often have less access to health services.
[93] This information is regularly available from standard Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS), so baseline data would be available. DHS and MICS are typically conducted every five years, but are staggered, i.e., data are available about every 2.5 years (thus monitoring for Gavi's PBF would incur additional costs if data are needed more frequently).

One cost-effective approach to verification would be to conduct small sample size surveys, which are cheaper than large sample size surveys. One example of such a survey is lot quality assurance sampling (LQAS), which is relatively simple to conduct and has been used by other major organizations working in global health, including WHO, the World Bank, and the Global Fund.

LQAS-based methods have been applied extensively in low- and middle-income countries to assess maternal and child health interventions, including in the assessment of vaccination campaigns.[94] Key advantages of LQAS are that the method requires a small sample size, is rapid, and as such comparatively inexpensive.[95] LQAS can be used to monitor program performance across a variety of indicators, including health outcomes and impacts as well as health system indicators.[96] It is also increasingly used to assess data accuracy,[97] and as such could be used to validate improvements in data if Gavi should want to reward such improvements as part of its PBF model.

LQAS surveys could be used to validate the data from countries on the selected performance indicators. However, they should be designed in a way that they also contribute to measuring the overall performance of the larger HSS grant, as the marginal cost for including additional measures (or indicators) to an existing LQAS survey will be minimal. This approach would prevent the creation of a standalone PBF verification mechanism - rather the LQAS surveys would become a valuable institutionalized mechanism to independently assess data quality and track changes over time. LQAS could also be used to measure the suggested equity indicator. While LQAS surveys come at a cost, which depends on the size of the survey and country context, they are still the most inexpensive

[94] Alberti, K. P., Guthmann, J. P., Fermon, F., Nargaye, K. D., & Grais, R. F. (2008): Use of Lot Quality Assurance Sampling (LQAS) to estimate vaccination coverage helps guide future vaccination efforts. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *102*(3), 251-254; Robertson, S. E., & Valadez, J. J. (2006): Global review of health care surveys using lot quality assurance sampling (LQAS), 1984–2004. *Social science & medicine*, *63*(6), 1648-1660. Valadez JJ. (1991): Assessing child survival programs in developing countries. Department of Population and International Health, Harvard School of Public Health. Boston, Massachusetts; Valadez JJ, Transgrud R, Mbugua M, Smith T. (1997): Assessing family planning service-delivery skills in Kenya. Stud Fam Plann. 143–50; Garner P, Smith GD. Information for decision making: Lot Quality Assurance Sampling in the spotlight. Int J Epidemiol. 2010; 39: 5–6.

[95] Robertson SE, Valadez JJ. (2006): Global review of health care surveys using lot quality assurance sampling (LQAS), 1984–2004. Soc Sci Med. 2006; 63: 1648–60. Valadez JJ, Devkota BR (2002): Decentralized supervision of community health program using LQAS in two districts of southern Nepal. Community-based health care: lessons from Bangladesh to Boston. Boston, MA: Management Sciences for Health.

[96] O'Connell, T. /Sharkey, A. (2013): Reaching Universal Health Coverage through District Health System Strengthening: Using a modified Tanahashi model sub-nationally to attain equitable and effective coverage, UNICEF. Pezzoli, Nick Andrews and Olivier Ronveaux (2010): Clustered lot quality assurance sampling to assess immunisation coverage: increasing rapidity and maintaining precision. Tropical Medicine and International Health doi:10.1111/j.1365-3156.2010.02482.x; Minetti et al. (2012): Performance of small cluster surveys and the clustered LQAS design to estimate local-level vaccination coverage in Mali. Emerging Themes in Epidemiology 2012, 9:6; Mwanza M, Zulu J, Topp SM, Musonda P, Mutale W, Chilengi R. (2017): Use of Lot quality assurance sampling surveys to evaluate community health worker performance in rural Zambia: a case of Luangwa district. *BMC Health Services Research*. 2017;17:279. doi:10.1186/s12913-017-2229-9.

[97] Gimbel, Sarah: Improving data quality across 3 sub-Saharan African countries using the Consolidated Framework for Implementation Research (CFIR): results from the African Health Initiative BMC Health Services Research 2017 17 (Suppl 3):828; Stewart JC, Schroeder DG, Marsh DR, Allhasane S, Kone D. (2001): Assessing a computerized routine health information system in Mali using LQAS. Health Policy Plan. 2001 Sep;16(3):248-55; Tikmani SS, Saleem S, McClure E, Naqvi FZ, Abrejo F, Soomro Z, Wallace D, Goldenberg RL. (2018): Monitoring of birth registry coverage and data quality utilizing lot quality assurance sampling methodology: A pilot study. J Family Med Prim Care. 2018 May-Jun;7(3):522-525. doi: 10.4103/jfmpc.jfmpc_59_17.

verification mechanism (it is true that the current mechanism comes at zero cost but our review has shown that it is not working).

**Recommendation 4: Gavi needs to significantly improve its communication to countries and Gavi staff about the PBF mechanism.** Our study found a need for better communication about what the PBF model is, how a country qualifies for it, and how it operates. Rather than intensifying communications on PBF immediately, we, however, suggest strengthening PBF communications when the revised PBF model is launched across countries.

***Learning recommendation***

**Recommendation 5: There needs to be better reporting of the results of Gavi's PBF and a stronger culture of learning.** The Secretariat should develop tools to effectively monitor the implementation and results of the PBF. These tools should leverage joint appraisals as an existing tool. Gavi also should check if any changes to the guidelines for applying for Gavi support might be required. Based on the data generated from these tools, it should report to the Board and the PPC on an annual basis to allow for discussion and learning. Progress will also depend heavily on learning from the best-performing countries in recent years, so Gavi should support South-to-South learning to ensure truly transformative shifts even in the poorest countries with the lowest coverage. This South-to South learning will also help to facilitate a more strategic and innovative use of PBF rewards across countries.