Scaling-up Results-Based Financing for Community Health

Learning Agenda Summarized Report: Key Results, Insights, and Recommendations

June 2023

LivingGoods   USAID FROM THE AMERICAN PEOPLE   GDI Global Development Incubator   INSTIGLIO TYING FUNDING TO IMPACT   DEERFIELD

Acknowledgements

This learning agenda was authored by the following Instiglio staff: Nicole Pflock, Tabitha Ngugi, Melissa Kaminker and Obed Matara. The report also benefited from the contributions and feedback of individuals from the following organizations: USAID DIV, Deerfield Foundation, Living Goods, GDI, and IPA. Their input was instrumental in shaping the insights and lessons highlighted in this report.

1. Introduction

1.1   About Living Goods

Since its founding in 2007, Living Goods (LG) has supported nearly 11,000 digitally empowered community health workers (CHWs) to deliver care, improving the ability of families to access the treatment and care they need.[1] LG has supported CHWs to go door-to-door in their communities, delivering an integrated package of reproductive, maternal, newborn, and child health interventions (see Figure 1).

LG supports its CHWs to deliver high impact community health services in a cost-effective manner by providing them with a smartphone and diagnostic health app, medicines and health tools, real-time supervision, and compensation for their work. LG commonly refers to this as the Digitally Enabled, Equipped, Supervised, and Compensation (DESC) framework.[2] LG also collaborates closely with the governments of Kenya, Uganda, and Burkina Faso to support them to develop their own CHW networks and build a policy environment needed to empower CHW in the long-term.[3] In addition, LG has been experimenting with innovative ways of financing community health programs to crowd-in additional funding for community health with the goal of sustainably reaching all communities in need.[4]

Figure 1. Services offered by LG CHWs

1.2   Scaling-up RBF for Community Health

Following the successful implementation of a one-year pilot Results-Based Financing (RBF) program in 2018, funded by Deerfield Foundation in two districts in Uganda (Masaka and Kyotera), LG decided to scale the RBF to three additional districts; Mafubira, Lira, and Wobulenzi, targeting approximately 2,000 CHWs. The RBF scale-up program, which began implementation in October 2020 for a duration of 27 months (9 quarters), aimed to **drive improved cost-effectiveness and quality of community health services targeting underserved and at-risk populations**. It also aimed to crowd in increased long-term funding for community health as well as support government engagement and the eventual uptake of RBF by the Government of Uganda.

Building on the lessons from the pilot, the design of the RBF scale-up program intended to place a stronger focus on the **cost-effectiveness** of CHW service delivery. The RBF attempted to drive efficiency in 3 key ways: i) by placing greater emphasis on improvements in CHW productivity, in other words driving efficiency through encouraging a greater number of visits conducted per each CHW instead of increasing results through the recruitment of more CHWs ; ii) increasing performance targets over time to encourage a culture of continuous improvements and, iii) implementing the RBF over a longer period of time to enable programmatic and operational adjustments. The RBF scale-up program also placed greater emphasis on the **quality of services** delivered by including in its design three metrics specifically focused on outcomes related to the quality CHW services delivered (client knowledge, client satisfaction, and CHW

---

[1] Capacity Statement Living Goods 2022: Living Goods supports digitally enables community health workers to save lives at scale
[2] Ibid.
[3] Ibid.
[4] 2022-2026 Strategic Plan. Saving Lives at Scale through Country-Led, Digitally-Enabled Community Health Systems.
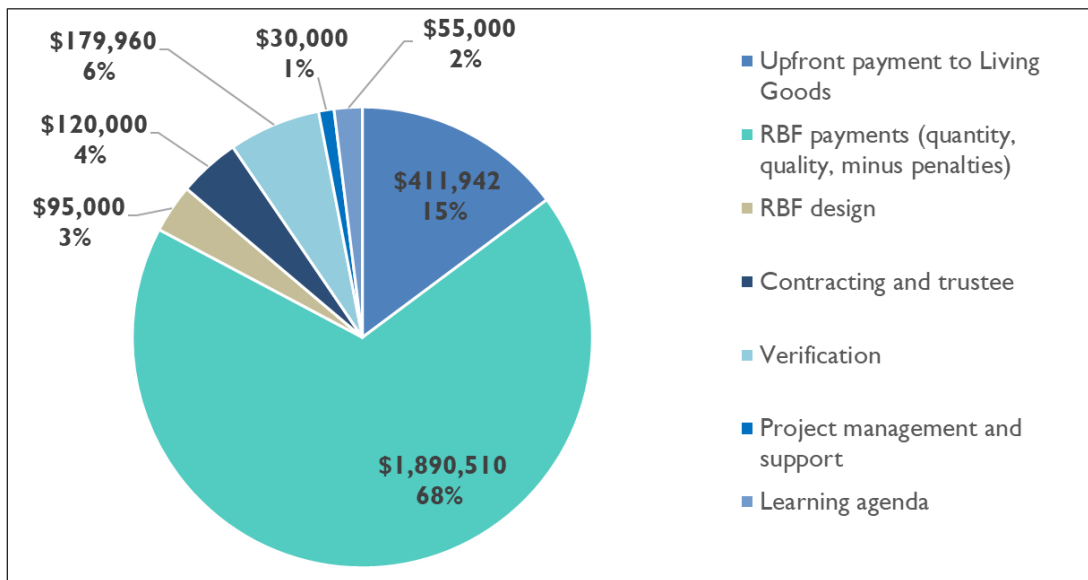
competence) and four quality safeguards, which incorporated penalties if performance dropped below a defined threshold (see Figure 2).

Figure 2. RBF payment metrics

Composition of payment

Quantity metrics
Maternal Health · Newborns · Children under 5 · Family Planning

90% of expected payment. Based on targets and prices, expected payment per health area were:
- Maternal Health + Newborns → 21%
- Children under 5 → 65%
- Family Planning → 14%

Quality metrics
Client Knowledge · Client Satisfaction · CHW Competence

10% of expected payments across the three metrics. Performance on quality metrics could increase payments earned on quantity metrics by up to 15% or decrease payments by up to 21% (sliding scale)

Quality safeguards
Unique households' coverage · CHW supervision · In-facility delivery · PNC visits

A penalty per safeguard of 5% if performance fell below a minimum threshold more than once

After an initial up-front payment, the scale-up program was financed by Living Goods with payments made post-fact for verified service delivery from an Outcomes Fund, with USAID DIV as the anchor outcome payer (see Figure 4). Over the duration of the program, LG also raised funds from Deerfield Foundation resulting in a total financial commitment for the Scale-up program of USD 2.8 million. **Of the total cost depicted in Figure 3 below, RBF payments constituted the largest portion,[5] representing 68% of the total cost.**

Figure 3. Total cost

- Upfront payment to Living Goods
- RBF payments (quantity, quality, minus penalties)
- RBF design
- Contracting and trustee
- Verification
- Project management and support
- Learning agenda

$179,960 6% · $30,000 1% · $55,000 2% · $411,942 15% · $120,000 4% · $95,000 3% · $1,890,510 68%

---

[5] The 15% upfront payment to Living Goods covered expenses related to program design, preparation, smart health app design configuration, government advocacy, engagement, additional outcome payer engagement.

Other stakeholders involved in the scale-up program included Global Development Incubator (GDI) who acted as the trustee; Innovations for Poverty (IPA) who was the Independent Verifier, and Instiglio who designed the RBF scale-up program and supported implementation as program manager (see Figure 4. RBF actor relationships).

Figure 4. RBF actor relationships
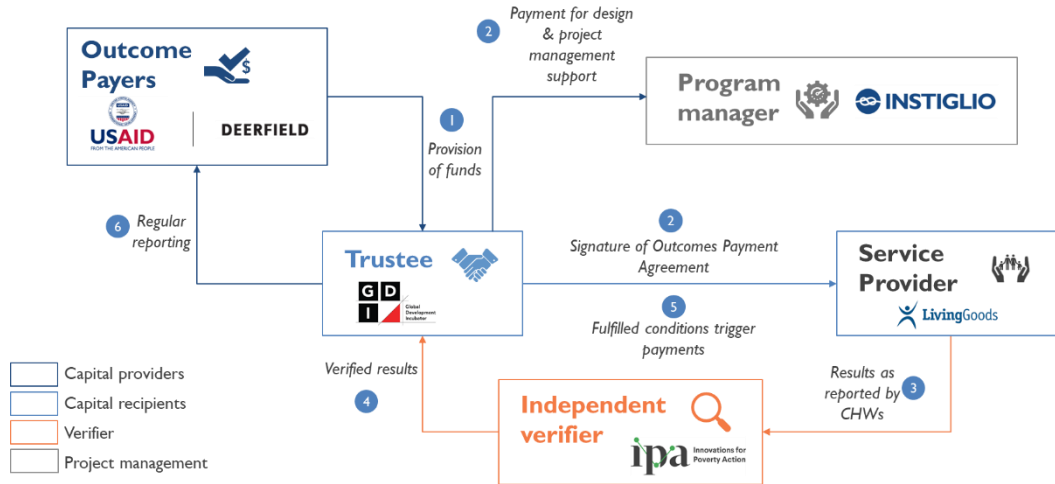
2. Key results from the scale-up program

On average, LG's CHWs exceeded targets on quality metrics[6] and attained 93% of the projected targets established for the quantity metrics.[7] Figure 5 shows average performance across different payment metrics. Although 93% indicates a strong and sustained level of performance, there was no evidence that the RBF motivated LG to improve CHW productivity (see Insight 1).

Figure 5. Performance snapshot

---
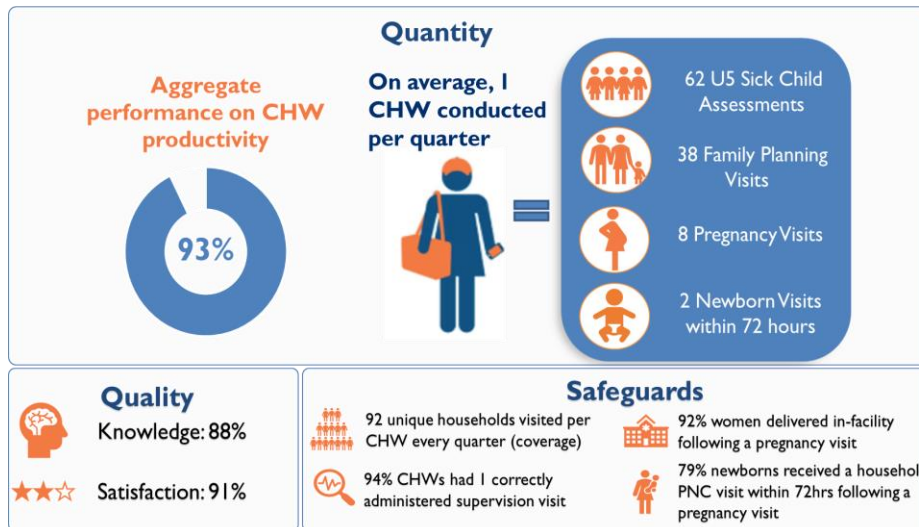
[6] The target for client knowledge was 70% and 75% for client satisfaction.

[7] The Learning Agenda assessed LG's performance based on CHW productivity as this was the main outcome the RBF intended to influence in order to drive cost-effectiveness. However, CHW productivity was only one factor that drove total outcome payments earned by LG. The other factor was number of active CHWs in the RBF program. As a result of LG not scaling to the total expected number of CHWs due to challenges raising additional outcome payments, LG earned only 67.9% of the total expected quantity payment (see Annex 1, Comparison to expected performance in the Learning Agenda Full Report for additional details).

3. Key insights

Insight 1. There is no evidence that the RBF led to improvements in CHW performance.

One of the objectives of the RBF scale-up program was to improve the cost-effectiveness of maternal and child health services delivered by LG. The main channel through which the RBF intended to drive cost-effectiveness was by increasing the number of results delivered per CHW, in other words, their productivity. While quantitative evidence indicated sustained performance on CHW productivity despite the COVID-19 pandemic with LG achieving 93% of the expected target on quantity metrics (see Figure 5) as well as 97% of historical performance, there was no indication that the RBF motivated an improvement in performance as a trend analysis that compared performance in RBF branches to comparable non-RBF branches found no significant difference between the two groups. While the latter could be indicative of positive spillover effects from RBF to non-RBF branches, qualitative evidence from interviews suggested that LG's primary focus was on understanding the reasons for the high verification error rate and addressing data quality concerns rather than improving productivity. Other factors that may have contributed to the reduced attention on productivity include:

1) **Stakeholders pursuing multiple objectives with the RBF**, including improving cost-effectiveness, testing new design features to promote improved quality of service delivery, attracting more funding from donors, and engaging government. This resulted in a relatively complex RBF design that included several features and drew attention to different factors. Additionally, the occurrence of COVID-19, and the shift in LG's long-term strategy, likely reduced LG's focus on and bandwidth to improve CHW productivity.
2) **Reporting performance** at the aggregate level (i.e., total results achieved by metric) rather than distinguishing whether underperformance was due to lower-than-expected productivity or fewer-than-expected CHWs, may have made it challenging for outcome payers to understand what was driving underperformance. While from Q5 onwards reporting was improved to highlight how performance was influenced by these two drivers, interviews suggest that this may not have been as impactful as the amount of information shared in verification and payment reports and quarterly review meetings may have been too much to absorb.

Insight 2. The RBF program led to measurable, scalable, and sustainable improvements in the quality of programmatic data on CHW performance.

The RBF's verification revealed significant data quality challenges, particularly during the first three quarters of the RBF with verification error rates of 49% (Q1) and 40% (Q2 and Q3) (see Figure 5).[8] The rigor of the independent verification approach – for example its approach to verifying metrics with a higher level of disaggregation (e.g., differentiating between U5 assessments and U5 referrals) and probing on specifics about a service provided (*see Section 3.6 in the Learning Agenda full report for additional details on the independent verification approach*) – enabled the depth of challenges to be revealed and encouraged a strong response from the highest levels of LG's management. Consequently, the independent verification drove LG to improve their Data Quality Assessment (DQA) guidelines and identify technological and process issues related to COVID-19 protocol changes and server capacity limitations. These changes led to measurable, scalable, and sustainable improvements in the quality of LG's programmatic data. Specific changes undertaken by LG included:

1) Developing and implementing a **data quality optimization plan** to ensure continued focus on data quality.
2) **Improving the training and capacity building of CHWs** to address knowledge gaps that were reported to be the main drivers of the poor quality of data.
3) **Modifying the LG Uganda program target-setting strategy** to reduce pressures that may have led to falsified reports. Interviews conducted as part of the Learning Agenda suggested that the previous strategy of setting targets at the programmatic level, without taking into consideration branch-level contexts, may have put CHWs under pressure to meet unrealistic targets, resulting in falsified reports. The shift to branch-specific targets that better

---

[8] While an error rate of 19% may still be considered high, there are several reasons that could be causing this other than misreporting by CHWs. These include (i) recall bias by respondents particularly as the independent verification had stricter conditions for a visit to be verified whereby enumerators not only probed on whether a visit took place, but also regarding specific services offered during the visit, (ii) incorrect responses by respondents especially if they found that the interview took too long or felt uncomfortable, and (iii) possible enumerator-led mistakes, e.g., based on how questions are framed.

reflected expected performance based on contextualized factors, such as disease burden, population coverage, seasonality of issues and historical performance, may have therefore served to improve data quality.

4) **Adopting stricter penalties at both CHW and supervisor level** to motivate a stronger focus on data quality.

5) **Updating SmartHealth app workflows** to address issues such as glitches in the Under 5 referral workflow which contributed to the high error rate by inaccurately prompting CHWs to make U5 referrals.

6) **Upgrading the SmartHealth App** and **revising protocols around syncing of data and hardware** use to minimize errors.

These changes not only improved data quality in the program districts (the error rate declined by approx. 60.4% between Q1 and Q9, see Figure 6), but they also proved to be **scalable and sustainable** in their applicability to non-RBF districts. They are also anticipated to **improve quality of service delivery** e.g., by ensuring CHWs always have access to a functional phone, LG will ensure that CHWs always have access to workflows to guide them in service provision.

Figure 6. RBF error rate

Insight 3. There is insufficient evidence to assess the RBF's contributions to the positive results observed on quality and safeguard metrics. This might be due to the inclusion of quality components in the verification of quantity metrics, which may have made it difficult to isolate the impact of the incentives on quality.

One of the RBF scale-up program's objectives was to test strategies to incentivize LG to maintain a high quality of service delivery. The RBF mechanism was designed to achieve this result by incentivizing performance on three quality metrics and applying penalties if a minimum threshold was violated on four quality safeguard indicators (see Figure 2). Quantitative evidence showed LG consistently exceeded targets on client knowledge and client satisfaction[9] (see Figure 7) with safeguard penalties triggered only once on the in-facility delivery safeguard indicator. [10]

---

[9] The analysis only focused on these two quality metrics as they were measured every quarter unlike CHW competence which was measured annually – at the end of Q4 and Q8 – and which LG and IPA were only able to report on/verify once due to delays in processing the Q8 recertification results. LG's performance on CHW competence for the first year was at 93%, an achievement of 98% of the expected target (95%).

[10] The in-facility delivery safeguard indicator was defined as percentage of women who delivered in a facility following a pregnancy visit. Every quarter, a penalty is applied if Living Goods' performance goes below the agreed threshold on a safeguard indicator and has received a warning in a previous quarter. LG had received a warning on the in-facility safeguard indicator in Q5 and a further violation of the safeguard's threshold in Q9 resulted in a 5% penalty being applied.

Figure 7. Performance on client knowledge and client satisfaction
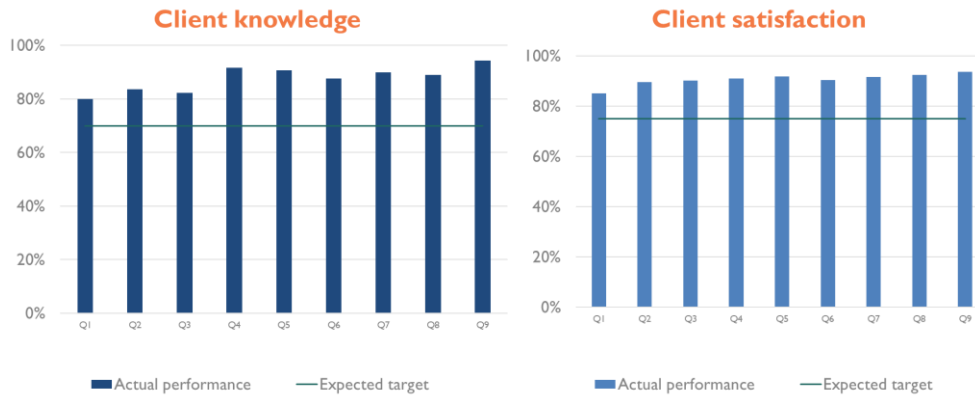
Despite the positive performance,[11] there was limited evidence on whether and how the RBF contributed to these results. This could be partly due to the verification of quantity metrics, which was dependent on evidence of quality service delivery. For example, a pregnancy visit was verified if the respondent confirmed the visit and could demonstrate 'client knowledge' by identifying one topic covered by the CHW during the visit, such as the benefits of going to a facility for an ANC visit. As quality service delivery was implicitly incentivized by the quantity metrics, it was challenging to establish whether the incentives placed on the quality metrics had an effect on performance.

As a result, a key recommendation for how to improve the measurement of quality metrics is to ensure that they are disentangled in the verification from quantity metrics. This would generate better performance insights into quantity and quality as separate aspects of service delivery and allow testing of different indicators and strategies to incentivize better quality of services.

Insight 4. Evidence suggests that most design features neither promoted nor constrained performance. Instead, design features ensured outcome payers only paid for verified results.

The scale-up program included nine key design features (see Figure 8). While all design features, excluding the quarterly payment cap and the renegotiation of targets, had an effect on LG's payment, qualitative and quantitative evidence suggests that only the verification and data quality penalty influenced a change in behavior and performance. On the other hand, there is no evidence to suggest that design features caused any perverse incentives (i.e., an overly strong focus on certain results). These findings suggest that the main contribution of the design features was to successfully mitigate the risk that outcome payers paid for results that were not achieved or were deemed less valuable during the design of the RBF or that LG would prioritize quantity over quality. For example, price kinks on the family planning metric were found to be effective in ensuring outcome payers did not overpay for performance on family planning.

---

[11] The performance on quality metrics show limited room for improvement.

Figure 8. RBF design features

1 Quarterly Payment caps
Every quarter, a payment cap, equivalent to 105% of the expected payment was included to encourage over performance and provide flexibility to make up for underperformance in previous quarters

2 Metric Specific Caps
Some metrics had caps to limit unnecessary visits and reduce the risk that relative prices were not well calibrated and would draw LG's attention to certain results

3 Price Kinks
Price kinks were included for U5 assessments and family planning visits to prevent unnecessary visits and provide incentives to perform well on all metrics

4 Relative Prices
Relative prices for quantity metrics aimed to align financial incentives with impact by paying a higher price for results with the greatest impact on maternal and child mortality

5 CHW Adjustment Factor
Provide incentives to achieve overall results through improvements in CHW productivity instead of adding additional CHWs (which would not lead to improvements in the cost-effectiveness)

6 Payment function linking quality and quantity
Aimed to incentivize a focus on quality-of-service delivery and mitigate the risk of neglecting service quality in favor of quantity

7 Safeguards
Safeguards were intended to ensure that a basic standard of quality was being met and that there would be a focus on coverage

8 Data quality penalty
Aimed to provide a stronger incentive for LG to improve the accuracy of its data

9 Renegotiation of targets
Provided LG an opportunity to request an adjustment of targets/prices in case of well-justified reasons (e.g., change in the disease burden) beyond its control, thereby ensuring a reasonable transfer of risks.

Insight 5. While the RBF mechanism was not designed or optimized for uptake by the Government of Uganda, the program did contribute to LG's advocacy efforts with the government by supporting to initiate engagements on potential strategies for improving the efficiency of their National Community Health Strategy

An important goal of the RBF scale-up program, noted during stakeholder interviews, was to test the potential for elements of the RBF model to be adopted by the Government of Uganda (GoU). However, a clear theory of change on how the scale-up program would influence government adoption was not developed and the RBF design was also not developed for adoption by government. Nonetheless, the RBF program was useful in supporting LG to initiate discussions with the GoU on innovative approaches to improving the efficiency of CHW programs. For example, LG's engagement strategy contributed to the government's inclusion of elements of LG's Digitally Enabled, Equipped, Supervised and Compensated (DESC) strategy into Uganda's first national community health strategy. This is evidenced most notably by one of the National Strategy's Strategic Objectives which directly references LG's approach in its goal to operationalize "a performance management framework for CHWs using the DESC approach."[12]

Insight 6. The design process of the RBF scale-up benefited from a robust RBF pilot design, lessons gathered through a process evaluation, and stronger capacity built within LG.

Specifically, the following factors contributed to an efficient scale-up design process:

1) The scale-up program built on a number of the RBF pilot design elements (e.g., payment metrics, targets, pricing methodology, verification methodology) and tools created (e.g., financial model).
2) Lessons and recommendations[13] gathered from the RBF pilot were used to refine the design of the scale-up program.
3) The scale-up program leveraged LG's organizational expertise and understanding of RBF, especially among key leadership, owing to their involvement in designing the pilot program and implementing it.

Despite being efficient, the integration of novel design elements, which included quality metrics, the addition of quantity metrics and expansion to new districts still required additional effort. For quality metrics, effort was dedicated to defining

---

[12] *National Community Health Worker Strategy (NCHWS), Strategic Objective 2.2.* Government of Uganda, Ministry of Health, February 2023
[13] Living Goods' RBF Pilot Internal Review; October, 2019.

the metrics and targets[14] and developing the verification protocol. For new districts, this included increased complexity in designing the financial model and determining prices due to variation in the performance levels across districts.

Insight 7. The RBF scale-up design was robust and adaptable to the COVID-19 pandemic, which did not trigger significant changes to the design.

The pandemic created uncertainty regarding LG's ability to meet performance targets because of disruptions to LG's operations that made the delivery of CHW services difficult if not impossible. For example, the pandemic was expected to divert CHW attention to the COVID-19 response, decrease productivity due to travel restrictions and illness or fear of infection among staff and CHWs, and delay LG's plan to scale-up the number of CHWs due to restrictions on gathering. To account for these risks, small modifications were made, such as (i) including an additional district (Wobulenzi) to account for challenges in scaling to the expected number of CHWs, (ii) providing LG the opportunity to renegotiate targets up to two times (instead of just once), between Q3 and Q7 and conditional on providing compelling reasons for adjusting targets,[15] and (iii) amending the verification protocol to conduct verification fully over the phone as opposed to a 50-50 split between in-person and phone verification.

Ultimately, the minimal adjustments made to the RBF design in response to the COVID-19 pandemic indicate that the design was quite robust and could allow LG to effectively adapt to the challenges posed by the crisis. It also reflects the stakeholders' commitment to find effective solutions that would uphold the program's objectives and spirit. This is evidenced by the fact that stakeholders opted not to have funds disbursed as a regular grant in the first two quarters – one option that was explored in the advent of COVID-19 – which would have reduced the time for testing and learning from the RBF. Instead, other solutions that still maintained the objectives of the RBF were explored and implemented. Crucial to note, however, is that although the adjustments made to the RBF were minimal, it did take extensive discussion to arrive at those conclusions and this burden may have somewhat offset the promised efficiencies of a classic RBF program.

The strong collaboration and commitment to problem-solving among stakeholders continued throughout the program. Stakeholders continued to openly discuss challenges and collaboratively find solutions, which was key to addressing other issues. For example, LG and IPA worked closely to both address challenges with the data sharing[16] process experienced at the start of the program (*see Section 3.10 in the Learning Agenda Full Report for additional details*).

4.   Lessons, reflections, and recommendations

4.1. RBF mechanisms can deliver value for money through their ability to accelerate learnings.

RBF mechanisms can function as incubators to enable organizations to **identify challenge areas, generate efficiencies and create positive spillover effects that can improve an organization's performance beyond a specific program.** For example, the RBF enabled LG to identify and address data quality challenges, which led to measurable, scalable, and sustainable improvements in the quality of programmatic data on CHW performance for RBF branches and non-RBF branches. In some instances, the challenges surfaced by the RBF mechanism were quick to identify and address. In other cases, the RBF pointed to where further effort and investment should be made to trace and address root causes. While addressing these challenges was costly for LG, the changes implemented – for example the development and adoption of a data quality optimization plan– may result in a positive effect on both productivity and quality of service delivery or reduce service delivery cost in the long-term.

On the other hand, the design, verification, and implementation of an RBF all require additional cost and staff time of all parties that should be factored into cost-effectiveness calculations. For example, the estimated additional costs for the design, verification, and implementation of the RBF during the scale-up amounted to USD 613,307[17] – despite the

---

[14] Determining appropriate targets was difficult due to insufficient historical data.

[15] LG did not renegotiate targets for the RBF.

[16] Initially, there was a lack of clarity regarding the technical requirements and process for sharing data for verification, leading to delays, incomplete or erroneous data. To overcome this challenge, a data generation and sharing guide was developed. In addition, a bi-weekly data sharing check-in was introduced to address any questions, concerns, or considerations related to the bi-weekly data shared by LG. These interventions led to a more streamlined and effective data sharing process between IPA and LG.

[17] Included in this estimate: RBF design (USD 95,000), verification (USD 179,960), contracting and trustee (USD 120,000), project management and support (USD 30,000) and USD188,347 for program design and preparation included in LG's upfront payments.

efficiencies leveraged from the initial pilot (see Figure 3). These costs may, however, decrease over time as efficiencies are gained, improvements and simplifications to the RBF processes are made, and economies of scale are realized.

4.2. The complexity of RBF mechanisms is contingent on stakeholder needs and objectives.

A key learning from the scale-up program is that when designing RBF mechanisms, stakeholders should carefully consider how their objectives could impact the design and consider approaching objectives in phases instead of simultaneously to avoid the need for a complex and thus expensive mechanism. It is especially important for the objectives of a program to be aligned with the capacity of the implementer and the maturity of a program.

This lesson is informed by the scale-up program, in which stakeholders pursued multiple objectives resulting in a relatively complex design. For example, the design included several design features to drive cost-effectiveness and enable testing and learning about how to incentivize quality-of-service delivery. During the design stage, the complexity was accepted given the experiences and lessons learned from the RBF pilot and LG's high capacity. During implementation, however, its complexity may have made it harder for LG to identify how to increase performance, particularly when navigating unexpected challenges, such as COVID-19 and the high verification error rates.

With regards to the maturity of the RBF mechanism, during the pilot and refinement phases of an RBF, the emphasis for stakeholders is typically on generating a broad base of learnings on how the mechanism functions, testing different modalities, while trying to mitigate risks comprehensively for multiple parties. These goals can drive up both the complexity and cost of a design. The focus of learning and evidence-generation may then become more targeted as a program matures and trust in programmatic data is strengthened, reducing the amount of rigor required and subsequently reducing the number of design features that an RBF needs to include. In such cases, more reliance can be put on complementary management strategies that are less costly to implement, for example relying on less frequent audits of data and data systems instead of resource-intensive verification approaches.

4.3. RBF mechanisms should seek in their design to balance risks of underpayment to the service provider with the risk of overpayment by the outcome funders.

While a key objective of RBF mechanisms is to mitigate against the risk of outcome payers overpaying for results, a crucial goal is to also ensure that service providers are paid fairly, i.e., that the risk of underpayment which is assumed by service providers, is mitigated. For this to be achieved, RBF designers and service providers should assess the potential likelihood and severity of unintended eventualities on the risk of underpayment compared to their incurred costs and ensure that they are factored into the design of the payment metrics and decision to accept this risk from the start. For example, in the case of the RBF scale-up program, when setting targets and prices, the RBF mechanism factored in the verification error rates observed during the RBF pilot. Additionally, LG was not penalized for error rates below 10% as a way of mitigating the severity of any unintended underpayment. If these factors were not taken into consideration and stakeholders insufficiently sensitized to them, resistance to utilizing an RBF mechanism may be created. It is crucial, therefore, that such design features are clearly communicated to all stakeholders and their function as a mitigation strategy to protect against underpayment to providers or overpayment by outcome payers is sufficiently understood and accepted.

4.4. Different strategies should be explored to improve the cost-effectiveness of RBF verification at scale.

One of the main costs of an RBF mechanism is often its verification mechanism, which is crucial for generating robust results evidence to calculate payments. In the case of the RBF scale-up program, the cost of verification accounted for 6% of the overall expenses of the program (see Figure 3). Inherent in the design of a verification mechanism is the trade-off that almost always has to be made between the cost of verification and its rigor. While stakeholders during the piloting, testing and refining phases of an RBF mechanism may be willing to invest more in verification in order to maximize learnings and meet the needs of low-risk appetites, these costs can become prohibitive as an RBF scales. As a result, identifying alternative effective strategies that reduce the cost of verification methodologies as a program matures is crucial for the sustainability of an RBF design. Some strategies that could be considered include:

1) Leveraging technology such as GPS mapping to confirm that CHWs visited a household[18] or locate survey respondents, which could reduce verification costs by minimizing the effort required to find respondents when conducting in-person verification.
2) Incentivizing CHWs to add and repeatedly verify phone numbers for all clients to allow for greater reliance on phone-verification which is less costly.
3) Auditing the implementor's verification approach (through system, process, and data assessments and/or random spot checks) to ensure it meets a minimum standard so that the implementors' own administrative data can be relied upon to calculate results payments. This can be an interesting strategy when implementers already have strong internal controls and data quality processes in place and are determined to be operating at a high capacity.

4.5. Understanding how to measure and incentivize the quality of performance is an area that requires further research.

One of the objectives of the RBF scale-up program was to learn about how to incentivize improvements in the quality-of-service delivery. However, the way in which the verification was structured generated limited evidence on whether it was specifically the incentives placed on the quality metrics – as opposed to the incentives on the quantity-focused metrics or other design features – that resulted in the strong performance on quality. To improve learnings on how the quality of services can be measured and incentivized, the following recommendations are made:

1) The verification of quality components of a program (how well a service was delivered) should be disentangled from the verification of quantity metrics (how often a service was delivered). This could mean, for example, that the verification of quantity metrics only considers whether a visit happened, while any details regarding what was discussed during the visit or client satisfaction would be assessed separately and not influence the analysis of the quantity metric. This may provide greater performance insights into quantity and quality as separate aspects of service delivery.
2) Invest in further research to identify ways to measure and incentivize quality of service delivery in RBF mechanisms that accurately reflects improvements while managing the risks of over- or under-paying a service provider. Most RBF approaches in global health focus payment metrics on the delivery of services (i.e., activities or outputs) as opposed to the outcomes of those services, such as mortality and morbidity, despite outcomes potentially being an important measure of the quality of services delivered. This is because it is often prohibitively costly, particularly at scale, to measure such outcomes. In addition, outcomes are usually further along in the results chain from specific CHW activities that it becomes challenging to confidently establish causation for any improvement in outcomes observed. For example, an increase in in-facility deliveries or vaccinations completed depends on an action to be taken by the person receiving care, which falls outside of the manageable control of CHWs. Placing a payment metric on that type of outcome as a way of incentivizing quality would therefore introduce a risk that the service provider may not be paid despite delivering a high-quality service. On the other hand, incentivizing outputs or activities also presents risks as the outcomes of health services arguably matter as much or more than the volume of health services delivered and should therefore also be considered. Figuring out this dichotomy is therefore critical for ensuring that an RBF can be effectively used and scaled.

4.6. Understanding how RBF governance structures can best establish clear protocols for decision-making while enabling flexibility and collaboration is an area that could benefit from additional research and development.

Stakeholders' commitment to learning and a governance structure that fostered collaboration and collective problem-solving was critical to the success of the RBF. However, too much flexibility can also lead to inefficiencies or undermine the integrity of the RBF by focusing stakeholders' attention on what adaptations of the RBF mechanism to make instead of the core objective of improving performance. To mitigate these risks, clear protocols should be established to identify which issues merit conversations on adapting the RBF design and who should approve which modifications.

While issues are often difficult to anticipate, stakeholders can and should use learnings from a 'pilot' or 'test/refine' phases of an RBF mechanism to inform protocols for at scale implementation when arguably, efficiency becomes more

---

[18] Funders may not consider this sufficiently reliable evidence to ascertain that quality visits were delivered

important. In addition, the sector would benefit from more best practices regarding governance structures and protocols that are efficient while not undermining flexibility/collaboration.