



Completed Impact Evaluations and Emerging Lessons from the Health Results Innovation Trust Fund Learning Portfolio

Eeshani Kandpal

Completed Impact Evaluations and Emerging Lessons from the Health Results Innovation Trust Fund Learning Portfolio

Results Based Financing (RBF) is an approach that is increasingly being utilized in various countries across different settings and levels to facilitate improvements to health system functionality and priority health outcomes. Since 2008, the Health Results Innovation Trust Fund (HRITF) has invested US \$420 million in HRITF linked to US \$2.4 billion in financing from the International Development Association to implement RBF programs in 32 countries. In many of these low- and middle-income nations, the Bank's board has also approved additional financing to expand these programs following the initial success of these programs. As well as sustaining these initiatives, the HRITF's learning portfolio evaluates the design, implementation, and impacts of the pilot programs funded by HRITF. Both quantitative and qualitative evaluations are utilized to assess, learn, and understand the impact of these programs.

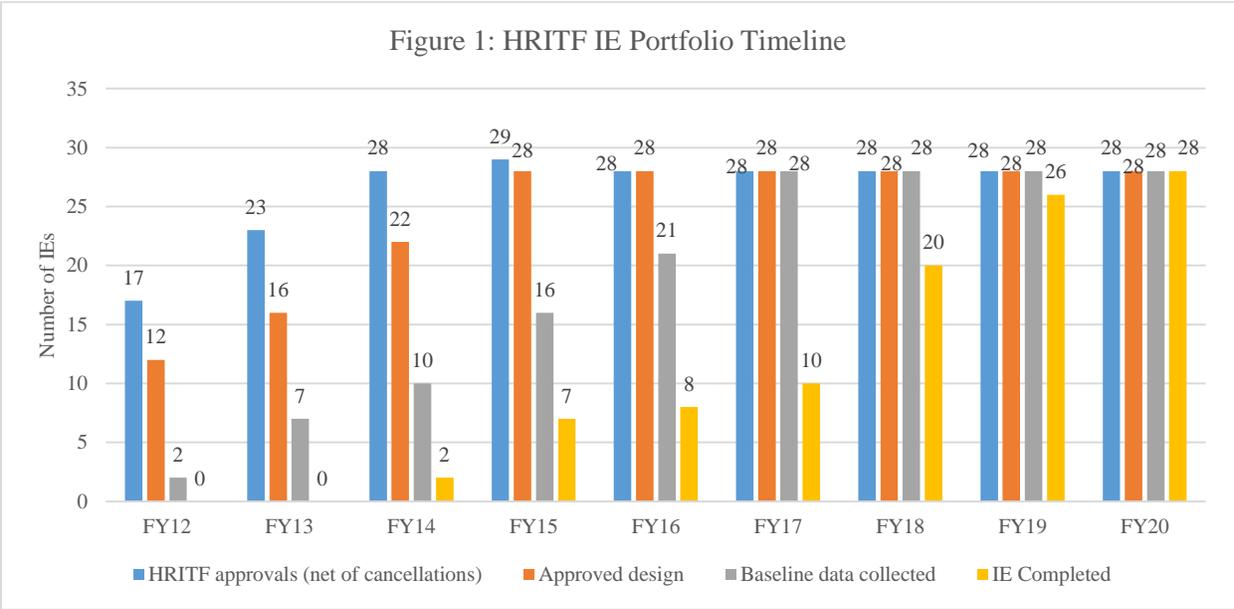
1. The evolution and composition of the HRITF Learning Portfolio

The HRITF learning portfolio comprises 29 rigorous impact evaluations, and 10 program assessments of such RBF programs focused on maternal and child health; these impact evaluations, and in particular the seven already-complete impact evaluations, are the focus of this summary note. Although not surveyed here, the HRITF learning portfolio also encompasses several additional learning activities, including case studies focused on learning from RBF implementation, which focus on learning from and during program implementation, including the analysis of administrative data, as well as analytical products which aim to learn on specific aspects of RBF, on topics such as family planning and incentives in RBF, cost-effectiveness, and qualitative tools to evaluate RBF. While evaluations are at different stages from early design to end-line analysis and dissemination, but an increasing number of evaluations of HRITF-funded pilot projects are reaching completion, providing evidence on varying aspects of RBF programs in maternal and child health. At the same time, several countries are also facing sustainability concerns as well as the decision to scale up.

Therefore, this is a pivotal time for the HRITF evaluation portfolio: several evaluations and learning activities have been completed that bring a critical mass of results that offer at times-divergent evidence on the effectiveness of RBF. These studies thus represent an opportunity to drive the learning agenda on RBF by both drawing on these completed evaluations and learning products, and fostering further learning amongst the evaluations and pilots as yet under progress. This note thus reviews the path of the HRITF-funded IE portfolio, focusing in particular on the results of seven complete comprehensive impact evaluations of RBF pilot projects, conducted in Afghanistan, Argentina, Cameroon, the district of Haut-

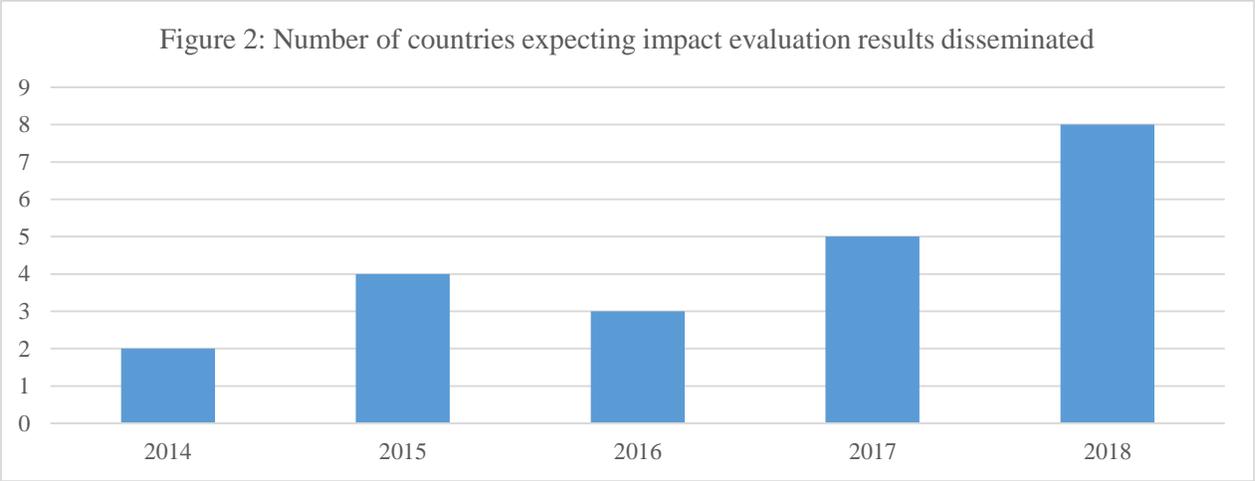
Katanga in the Democratic Republic of Congo, Rwanda, Zambia, and Zimbabwe. In addition to taking stock of the HRITF portfolio, this study seeks to provide some new insights on whether RBF can be effective in improving health systems and health outcomes by: (1) synthesizing results of key completed IEs, (2) identifying key lessons that can be drawn jointly from these IEs, and (3) identifying crucial questions that are yet to be answered about RBF programs in health service delivery. The goal of this note is thus two-fold: (1) to take stock of the HRITF’s learning strategy as it stands today, and (2) to start building the evidence base to inform policy-making on results-based financing in health.

Over the past five years, the HRITF has built capacity for impact evaluations amongst country teams and teams are progressing towards producing results. Figure 1 shows the growth of the IE portfolio. Even as recently as FY12, there were only 17 IEs with approved funding, while the target of 28 IEs has already been achieved by FY15¹. By FY16, 7 IEs had reached completion, and it is anticipated two more will each completion in FY17 for a total nine completed IEs and one program assessment. Another sign that the IE portfolio is on track to generating a rigorous evidence base on RBF by FY20 is that all of the IEs will have completed baseline data collection by the end of this fiscal year. As a result, there will be at least 18 months to 24 months for exposure in each IE, which will allow for comparability as well as time for analysis by FY20. This chart also demonstrates how, even though prospective impact evaluations can take several years to completion, the IE portfolio is regularly generating new evidence on the effectiveness of RBF for health.



¹ A World Bank Fiscal Year (FY) starts July 1 of one year and ends June 30 of the following year.

Figure 2 highlights the expected timeline of results from the impact evaluations (excluding the program assessment and enhanced program assessments). While new results are expected every year, fifteen impact evaluations² remain at the early design stage, for which the timeframe still is to be confirmed, and also for which there are opportunities to drive the learning agenda on RBF more proactively. In addition, many IEs in the HRITF portfolio use a combination of quantitative and qualitative tools to help us understand the mechanisms behind any observed impact, and to unpack the “black box” of complex RBF projects into the specific program design elements.



The HRITF has achieved the objective to facilitate learning on RBF through the development and use of a conceptual framework that disentangles the causal chain according to which one would expect RBF to produce positive results on healthcare delivery, quality of care, healthcare utilization and ultimately, health outcomes. This conceptual framework currently pertains to Performance-Based Financing(PBF) schemes to respond to the immediate demand for learning in the portfolio, however it can easily be adapted to other types of RBF and includes key behavioral and organizational changes that RBF, broadly speaking, is to bring about. The current learning activities of the HRITF are embedded in that framework and the current gaps of the HRITF learning portfolio are mapped out according to that framework.

In addition, over time, HRITF has identified learning gaps and initiated learning activities focusing on key aspects of RBF. The analytical product on verification in RBF has been finalized and reported on to donors, while several case studies highlighting the current verification system of various countries, its strengths and gaps, have been published on the rbfhealth.org website. The human resources for health

² These impact evaluations are being conducted in Armenia, Benin, Burkina Faso, China, The Democratic Republic of Congo, The Gambia, Ghana, Haiti, Kyrgyz Republic, Liberia, Nigeria, The Republic of Congo, Senegal, Tajikistan, and Zimbabwe.

study has also been finalized recently; findings were disseminated through a blog post and paper. Through these analytical products, impact evaluations, LFRI case studies, the HRITF has built a comprehensive approach to learning, using a variety of methods to learn on the diversity of RBF in a timely manner.

Through the active encouragement of impact evaluations, whereby a rigorous impact evaluation generally has to accompany each Country Pilot Grant, the HRITF cultivates a culture of evaluation and learning within country teams implementing RBF projects. Over time, several country teams came back to the HRITF with more ideas and topics to learn about. Knowledge dissemination through the RBF website and the sharing of knowledge and experiences through seminars and workshops that bring the various implementing countries together, play a pivotal role in the HRITF program. These activities help enhance the learning and application of this learning to countries implementing and evaluation RBF programs.

The HRITF provides all needed technical support to country teams to ensure that data are promptly and carefully analysed and that all technical reports meet pre-established quality standards. These reports will be carefully reviewed by the HRITF team to ensure ~~that~~ the quality of the analysis. Timely dissemination will be encouraged and supported. The HRITF team also supports the country teams to make the data and documentation publicly available in a timely manner. The synthesis of lessons from the HRITF portfolio is an ongoing process and will strive to incorporate new results, judiciously incorporating important new lessons and nuances as the second generation of IEs provides a wide-ranging set of lessons on topics including purchasing quality. The timely dissemination of results is another key component of the knowledge built by the HRITF portfolio. To this end, the midterm review of projects receives inputs from the associated impact evaluation, and in many cases such midterm reviews have incorporated course corrections based on the input from the IE. Results are, of course, first disseminated in country, and then more widely through the IE workshops, health economics and policy conferences and other seminars. Furthermore, the team will ensure that impact evaluations with the more complex lessons or a wider reference population are actively sharing these lessons with the HRITF community as well as the wider RBF community of practice.

In addition, the HRITF conducted several training and technical assistance activities which helped build team capacity on impact evaluation and LFRI. For instance, regularly-conducted workshops bring together the country teams implementing and evaluating RBF projects to share findings and to learn from each other. In addition, targeted workshops have been held for several country teams, including Benin,

Cameroon, Central African Republic, Nigeria and Rwanda. During these workshops, teams have worked with HRITF staff to work on IE design and research question selection, data analysis and report writing and dissemination. By building on this culture of evaluation and creating local ownership of the IE and its results, HRITF views these capacity building activities as a vital component of its efforts to build evidence on RBF programs.

2. Summary of key results thus far

Seven completed impact evaluations in the HRITF portfolio provide key evidence on the impact of RBF programs in health service provision. These impact evaluations were conducted in Afghanistan, Argentina, Cameroon, the Haut-Katanga district of the Democratic Republic of Congo, Rwanda, Zambia, and Zimbabwe. For each of these countries, this note first reviews any notable results on (1) coverage indicators, (2) measures of structural and process quality³, (3) human resources, and (4) cost-effectiveness.

Afghanistan implemented a third party-implemented P4P pilot from 2010 to 2015 at two levels: primary care and hospital. In both cases, facilities are provided a performance bonus of up to ten percent of the value of their existing contract with the Government based on a quantity and quality checklist. Additional quality-based payments were made to hospitals but not primary care facilities.

The evidence from Afghanistan comes from two separate endline studies. Due to security reasons, the initial impact evaluation, conducted by Johns Hopkins University, had to be cut short, and the evaluation team had to move the endline survey up in time, thereby substantially shortening the exposure period, and also leave out two provinces due to security reasons. Therefore, in order to gauge the impact of the intended exposure period and the inclusion of all treated provinces, a second endline survey was conducted in all the treated provinces after the security situation stabilized. We discuss results from both endline surveys. In the P4P pilot in Afghanistan, randomly selected primary care facilities from matched pairs of facilities were assigned either to the P4P arm or the control arm in 11 provinces. P4P facilities were given bonus payments based on the volume of nine health services provided. Additional annual payments were made based on a balanced scorecard assessing the quality of services, and contraceptive prevalence rates to the implementing NGOs.

³ Structural quality refers to the hospital buildings, financing, equipment, supplies, and staff. On the other hand, process signifies all the interactions between patients and providers throughout the delivery of care.

The first evaluation found no significant differences in any of the five MCH coverage indicators (modern contraception, antenatal care, skilled birth attendance, postnatal care, and childhood pentavalent vaccination) or indeed on equity measures between the P4P arm and the comparison arm. However, substantial increases were observed in the quality of patient examinations and counselling, as well as in the time spent with patients. The study also noted significant difficulties communicating the intervention design to the health workers, and that health workers reported not understanding the structure of the bonuses. Results from the second endline are largely similar, with few statistically significant differences between the treatment and control arms on utilization of health services. However, focusing on the nine provinces surveyed in the first endline as well, which remained relatively stable throughout the program's duration, the study found improvements in staff training and the availability of medical supplies, as well as in facility-level management. Qualitative work showed that the RBF program was considered a good additional motivator even though salaries and the incentive were not always received on time.

Starting in 2004, Argentina's *Plan Nacer* delivers insurance for MCH services to uninsured families. The program allocates province-level funding on the basis on beneficiary enrollment as well as providing incentives following a pay-for-performance model based on indicators of the use and quality of MCH services and health outcomes. The provinces use these resources to pay health facilities to provide maternal and child health care services to beneficiaries. Two studies provide evidence on this program—the first was an impact evaluation of the entire program that used civil registration data on birth records in seven Argentine provinces for 2004 to 2008, while the second uses administrative data from the Misiones province to test the long-term impact of a 2010 experiment that temporarily but steeply increased incentives paid to providers for increasing early initiation on perinatal care.

The *Plan Nacer* IE found that the program increases the number of perinatal care visits as well as the probability of receiving a tetanus vaccine. Beneficiaries also had a 19% lower probability of low birth-weight compared to non-beneficiaries. Beneficiaries have a 74% lower chance of in-hospital neonatal mortality in larger facilities, approximately half of which reduction is due to from fewer low birth weight babies and half from better postnatal care. Results also show that the financial autonomy provided to facilities by *Plan Nacer* allowed a better allocation of scarce resources, which in turns had positive impact on health outcomes of beneficiaries. The cost-effectiveness analysis finds *Plan Nacer* to be highly cost-effective compared with Argentina's gross domestic product per capita over this period. However, the

study also found small negative spillover effects on prenatal care utilization of non-beneficiary populations in clinics covered by Plan Nacer, but no spillover was found on birth outcomes.

The second evaluation of the Argentinian RBF project focuses on the Misiones province and uses a randomized field experiment to provide key evidence on the sustainability of effects of RBF incentives. The evaluation estimates the effect of large (three-fold) but temporary increase in financial incentives for health care providers on the initiation of prenatal care in the first trimester of pregnancy. Results show the rate of early initiation of prenatal care was 34% higher in the treatment group than in the comparison group while the incentives were being paid, and that this effect persisted 12 months after the incentives ended. Results, however, also suggest that the quality of care may have remained a constraint to improving health outcomes as the increase in early initiation of prenatal care did not have any effect on birth outcomes. Nonetheless, the study also finds that large-but-temporary incentives can be more cost-effective at motivating provider performance and changing clinical practice than permanent incentives when providers face fixed costs to changing clinical practice routines.

Cameroon started a PBF pilot in 2012 in 14 districts in three regions of the country. The evaluation compared four arms: (1) the standard PBF package, (2) the same level of financing but not linked to performance, and with the same levels of supervision, monitoring, and autonomy as PBF, (3) no additional resources or autonomy, but the same levels of supervision and monitoring as PBF, and (4) pure comparison. For the PBF arm (arm 1), preliminary results show significant increases in coverage of the children vaccinations (including the polio 3 vaccine) and maternal immunization against tetanus as well as the coverage of modern methods of family planning, but no significant changes for timely ANC and in-facility deliveries. In terms of structural quality, the PBF arm (arm 1) as well as the additional financing arm (arm 2) saw increases in the average availability of necessary equipment, particularly materials for delivery and neonatal care. Process quality also appears to have improved in the PBF arm, with more qualified health workers present on site than in the other arms. Even though out-of-pocket health expenditures decreased for household in the PBF arm, this decrease in revenue did not come at the cost of process quality: there were no negative spillover effects on completeness of services and advice provided during antenatal visits and consultations for children under 5. Perhaps not surprisingly, then, given these impacts on quality of care, client satisfaction also increased for medical consultations for children younger than 5.

The Haut-Katanga district of the Democratic Republic of Congo started a P4P pilot in 2009. In the DRC pilot, facilities' payment was determined by the quantity of services provided relative to the other health

facilities rather than to the quality of care provided. In contrast, the amount allocated to each facility in the comparison group was calculated based on the staff in the facility. Health facilities were randomly assigned to either the treatment or comparison group. Both groups benefitted similarly from construction/rehabilitation, provision of equipment and medicine, staff training, and user fee reduction policy. However, payments in the treatment group were supposed to be contingent on external verification of results, while payments in the comparison group were solely based on the number of rank-specific staff employed. In addition, in the treatment group health facility managers were granted increased financial autonomy to distribute the payment subsidies among facility staff. However, the implementation of the program deviated significantly from the intended design of the program. Most significantly, there were several interruptions in the payment schedule and the comparison group received substantially higher payment levels during the first six months of the two year-long intervention. Furthermore, as facilities in the treatment group reduced the prices of their services but were unable to attract more patients, revenues were substantially lower in the treatment group than the comparison group, leading to 42% fewer resources for the treatment facilities, and 34% less income for health workers in the treatment group than in the comparison group.

Despite these challenges, the evaluation finds that financial incentives linked to performance outcomes led to a 43% increase in the provision of preventive sessions for targeted services in the treatment group than in the comparison group, with no discernible impact on the quality of these services or associated patient satisfaction. In addition, the autonomy associated with financial incentive system in the treatment group also led to a more egalitarian distribution of the bonus payments among facility staff. However, these increases did not translate into higher levels of service utilization or better health outcomes. While the evaluation questions the sustainability of RBF programs because the RBF arm was associated with significantly lower intrinsic motivation after the intervention ended, such de-motivational effects may equally likely have been caused by the significantly lower resource availability in the treated facility.

In 2010, the 'second generation' of PBF in Rwanda, the Community Performance-Based Financing (CPBF) program, was initiated to tackle the remaining issue of low utilization of health services by mothers and their children. The Community PBF implemented the following three interventions: (i) demand-side in-kind incentives for women, (ii) performance-based payment for community health worker (CHW) cooperatives, and (iii) combined demand-side and CHW cooperative performance payments. The three CPBF interventions were introduced in October 2010; the evaluation randomly allocated sub-districts of

the country into four study groups: three implementing the above CPBF interventions, and one pure comparison group. The first PBF pilot in Rwanda, started in 2005, provided bonus payments for the quality and quantity of 14 targeted MCH indicators directly to primary health centers in all districts without a pre-existing pay-for-performance scheme. The evidence from this study showed a large and significant positive impact on institutional deliveries, the quality of prenatal care, and the utilization of preventive care for young children.

During this period, Rwanda independently experienced significant and continuous progress made with respect to many indicators of maternal and child health. Nonetheless, comparing baseline and endline outcomes of the comparison group that was not exposed to the demand- or supply-side incentives schemes introduced by CPBF, the IE finds that the supply-side incentives to CHW cooperatives did not affect any of the outcome indicators, with no significant differences in the rate of women who report timely ANC and PNC or skilled-attended in-facility delivery or in self-reported behaviors of CHWs including the number of hours spent on health work, the number of households visited or the frequency of consulting other CHWs. CHW satisfaction and motivation did not change overall, although there is little variation in the responses to these questions. The demand-side in-kind incentives, on the other hand, are found to have a significant positive impact on timely antenatal and postnatal care. Relative to the comparison group, women in the demand-side-intervention-only arm were almost 10 percentage points more likely to attend ANC during the first four months of their last pregnancy and 7 percentage points more likely to attend PNC within the ten days after delivery.

However, the data also show some challenges with the implementation of the supplementary demand-side program of conditional in-kind transfers, with many health centers reporting difficulties with the procurement of the gifts and experiencing frequent stock outs. To some extent, these reports might be due to the fact the endline survey took place months after the last funds were transferred to the health centers. Among the women who reported being eligible for the transfers during the years in which the program was implemented in their sectors, a majority reported not receiving the transfers. On the other hand, preliminary qualitative evidence from Rwanda also suggest that the involvement of the CHWs was key to the program's success as they helped create demand for health services by engaging with the communities.

In 2012, Zambia implemented an RBF pilot project in ten districts. The Zambia IE was a three-arm evaluation that tested RBF against an enhanced financing-only arm and a pure comparison arm. Of the nine indicators directly targeted by the RBF program through the incentive structure, some responded to

the RBF program, with a broadly similar set also showing improvements under the enhanced financing arm. Institutional deliveries and skilled birth attendance increased substantially in the RBF districts relative to the comparison arm, but rose by even more in the enhanced financing arm relative to the pure comparison arm. One of the most important gains in the RBF arm was that the first ANC visit was earlier by two weeks as compared to the two other arms. In addition, while full vaccination coverage declined in both, the enhanced financing and comparison arms, it remained constant or slightly higher in RBF districts, suggesting that the RBF program was protective with respect to some measures of immunization coverage.

On the other hand, the rate of change for any PNC was more rapid in the enhanced financing arm as compared to RBF districts.

The evaluation also found some improvements in structural quality in the RBF arm and mixed results on process quality: patient counseling appearing to have worked better in the RBF arm, but testing and supplement provision being better in the enhanced financing arm. Health workers in RBF facilities also spent significantly more time during consultations with their patients as compared to both other arms. In contrast to the DRC and Rwanda pilots, the Zambia study found large gains in health worker satisfaction and staff motivation. Results show that health workers received about 10% of their official staff salaries on average as RBF staff incentives. It is important to note that by the end of the program, the proportion of disbursement to the enhanced financing arm as compared to the RBF districts was only 56%. Lastly, results from the cost-effectiveness analysis show that the RBF program is cost-effective in improving maternal and child health when compared to the pure comparison arm, but the results are mixed when compared to the enhanced-financing arm; nonetheless, the estimates generally point to favorable impact of the RBF arm.

The Zimbabwe RBF pilot program has three components: (i) results-based contracting; (ii) management and capacity building; and (iii) monitoring and documentation. Unlike in the Haut-Katanga, where quality was not incentivized, under the first component of the Zimbabwe pilot, a portion of financing received by health facilities depends on the quantity and quality of services, with a focus on maternal and child health. User fees were abolished in RBF pilot program districts, with the aim of improving access to care. Key indicators such as delivery by skilled provider and in-facility delivery improved faster in RBF districts than in control districts. Findings suggest a largely positive message around quality with *some* dimensions showing significant improvements under RBF, but not others: improvements were observed for selected measures of structural quality, as well as increased availability of certain equipment. Conversely, for both

the RBF and control districts, the availability of the majority of medicines, supplies, and equipment remained relatively unchanged, with minor fluctuations across products from baseline. A key indicator of the system-level effects was the strong evidence suggesting no neglect of non-incentivized services. Of note, none of the non-incentivized services investigated showed a decline in the number of cases treated, as would be expected if task shifting affected these services. For many of these services there appears to be a slight increase in volume, suggesting that service coverage actually increased for a broader set of services than those directly incentivized. Key enabling factors included improved autonomy, decentralized decision-making, and strengthened facility-level governance. Findings also suggest greater positive effects for the lesser-educated groups and the poor. Finally, the cost-effectiveness analysis finds the intervention to be highly cost effective; indeed, it is as cost-effective as a single-purpose MCH intervention, even without accounting for broader health system benefits.

However, as in the Haut-Katanga pilot, the RBF program had varied effects on health worker motivation. According to the qualitative findings, although staff were strongly motivated by incentives and their improved ability to serve the community, they also expressed their dissatisfaction with reduced service unit prices, the proportion of incentives relative to their tasks and those of their peers, limited leadership ability among supervisors and heads, and increased patient load, contributing to a higher workload and consequent burnout.

3. Summary of key lessons thus far

Even with a relatively small number of complete IEs, the HRITF portfolio already provides indicative evidence suggesting that RBF approaches to health delivery can be highly effective and even cost effective when it comes to improving coverage and the quality of targeted services across many aspects of maternal and neonatal health. While this evidence is by no means definitive of the global impact of RBF in maternal and child health, it comes from rigorous impact evaluations in Argentina, Cameroon, Rwanda, Zimbabwe, and Zambia, and to a limited extent, from the Haut-Katanga pilot in the DRC. There is also evidence of general health system strengthening in terms of more active supervising and monitoring roles, more quantifiable involvement with communities, and increased health worker satisfaction. These system-level impacts can have knock on effects on population-level health outcomes that may extend well beyond the life of the evaluation period.

Methodologically, the HRITF IEs have been encouraged to use cost-effectiveness studies, discussed in depth in the synthesis of the qualitative learning from the HRITF portfolio, and qualitative methods to complement quantitative results. Indeed, such studies have been relatively successful at unpacking the

various components that typically comprise an RBF intervention in Afghanistan, Cameroon, Rwanda, Zimbabwe and Zambia. However, program process monitoring and evaluation must also be used systematically to enhance learning from implementation, which is critical for the successful implementation of an RBF program. Mid-course changes effected in the Zimbabwe program were largely informed by the process monitoring and evaluation studies jointly commissioned by the government and the Bank.

Evidence, both from within the HRITF portfolio and from external evaluations, illustrates that demand-side and supply-side incentives work on different margins and may work best when combined—demand-side incentives can directly increase health-seeking behavior beyond what supply-side incentives can achieve by themselves. Improved quality and outreach by health providers can have an additional, indirect impact. In addition, the early evidence from the HRITF IE portfolio suggests that supply-side incentives encourage health providers to deliver more and better care to people who have made it to the facility. However, given the complexity of the RBF interventions, it is also important to understand the nature of the bottlenecks to improving health outcomes. For instance, the qualitative work in Afghanistan identified the lack of attention to demand-side considerations as one of the flaws of the RBF pilot implemented there. Such consideration of demand-side aspects need not necessarily come in the form of a voucher or cash transfer: early results from the qualitative study in Rwanda also suggest that the involvement of the CHWs was key to the program’s success as they helped create demand for health services by engaging with the communities. Relatedly, evidence from Zambia and Cameroon points out that, depending on the targets, enhanced financing can also produce good results on indicators as shown in this study. This is an area that would benefit from further research and testing. One of the first evaluations of the RBF mechanism in Rwanda found strong gains in targeted indicators against a compensated financing arm, but was unable to evaluate the enhanced financing arm against a pure control.

Another common theme in the results from Argentina, Afghanistan, Cameroon, Zambia and Zimbabwe is that RBF schemes effectively improve autonomy at the facility level. Even the evaluation of the first generation RBF pilot in Rwanda noted that provider-control is key to success. The autonomy associated with RBF enables more responsiveness to health facility needs by health workers, which not only benefits incentivized indicators but perhaps also enables health providers to address broader health systems challenges such as stock-outs for drugs and non-incentivized conditions such as non-communicable diseases, as exhibited by increased reported services in these areas. A better allocation of scarce

resources has a positive impact on health outcomes. Direct disbursement of funds to front-line service delivery levels, an effective disbursement mechanism, and decentralization of expenditures can further increase predictability of funding and managerial autonomy.

RBF programs also had significant and generally positive effects on the quality of care provided: the IEs from Afghanistan, Cameroon and Zimbabwe all show measurable improvements in structural and process quality. In addition, the IE in Zambia found some improvements in structural quality in the RBF arm, and somewhat more limited, but positive, results on process quality. Health workers in RBF facilities in Afghanistan and Zambia also spent significantly more time during consultations with their patients. While another concern may have been that incentivizing certain services led to a decrease in the quality or quantity of non-incentivized services, the DRC pilot shows that, despite an increase in the provision of preventive services, the quality of care did not change, suggesting the absence of spillovers from the provision of more services. The IE from Zimbabwe shows that RBF facilities also provided more non-incentivized services without measurable changes in process quality.

While the majority of impacts on quality appear to be positive, the results on human resources for health outcomes such as health worker satisfaction and motivation and evidence from the qualitative studies point out that (i) uncompensated price reductions of RBF services can induce negative effects in motivation among health workers, as in Zambia and the DRC pilot; (ii) the need to start at a sustainable level in pricing structures of incentive schemes and introduce increases based on robust financial analysis, as in Zimbabwe; and (iii) team incentives play a positive role in health facilities beyond just monetary transfers as in Zimbabwe.

The evaluations reviewed above also demonstrate the importance of continued innovations on ways to intelligently measure and incentivize quality measures of care in maternal and child health, which are more complex than coverage indicators. Related to this, given that quality of care is multidimensional, starting with structural quality indicators and then progressively introducing process measures of clinical care is critical to allow health providers to address less complex quality of care issues first, develop better understanding of RBF and quality of care, and then shift gradually toward more demanding measures of care under the RBF pilot.

While this finding may at first glance appear obvious, adequate levels of incentives need be offered to providers to affect behavioral change. The relatively low power of RBF incentives in relation to guaranteed salary (often on the order of 10% of salary) may limit some of the possible gains from RBF schemes. Indeed, the strongest evidence for sustained impacts from RBF comes from the Misiones

province, where the increase in incentives was substantial—threefold. It may also be the case that the signaling effect of an incentive introduced by a health system in an environment that previously did not incentivize individual services may be somewhat more effective at changing behavior than the income effect of the relatively small incentive amount offered. As such, provider effort may be relatively inelastic at small incentive levels apart from the “nudge” of the introduction of the incentive scheme. Further RBF studies from other settings will hopefully shed further light on this issue. On the other hand, the varying impacts estimated by these evaluations suggest the importance of continuing to study the effectiveness of RBF programs in new contexts before scaling up the intervention.

The early evidence also highlights that RBF mechanisms are not always easy to implement and have been associated with implementation failures that result in less effective programs; in a pilot conducted in the Haut-Katanga district of the Democratic Republic of Congo problems in implementation resulted in significant decreases in pay for health workers in the RBF facilities, leading to an unsurprising loss of motivation among these health workers. As noted by the impact evaluation in Afghanistan, the complexity of the RBF approach, and indeed of the payment formula itself can impede understanding of the incentive mechanism, which in turn prevents an adequate behavioral response to the incentive-based payment on the health worker’s part.

In addition to the importance of implementation, lessons learned also emphasize the readiness of the health system for a relatively complex health system reform. Other lessons include the judicious use of financing incentives as spending efficiency may not be high if certain targets are universally or near-universally provided: across all the evaluations reviewed here, some of the coverage indicators that exhibited the lowest degree of change under the RBF also exhibited the highest baseline coverage rates, suggesting that incentivizing these indicators may not present a highly efficient leverage of program funds. For national health systems with high coverage indicators, target or coverage based performance incentive frameworks might be better suited rather than fee-for-service. Yet another alternative to be considered might be flexible fee-based weights that can be modified to indicate current health system priorities. In summary, careful thought should be given to selected indicators in future program design in order to maximize the efficiency of spending.

Nonetheless, the HRITF IE portfolio remains a young one, and despite the substantial investment in RBF-based reforms of health systems, the rigorous evidence base remains sparse. The results thus far strongly suggest that RBF should not be isolated from broader health systems reforms and complementary interventions. Instead, RBF should be viewed as an entry point to tackling wider systemic issues that are

brought to the fore when RBF is rolled out. A good example is seen in the lack of demand for health services in Afghanistan and the DRC, and in the human resources for health management and coordination challenges at health provider level in Zimbabwe: these could greatly benefit from parallel health sector reforms to increase demand and strengthen health facility management and accountability.

As the HRITF continues to generate evidence on the effectiveness of RBF programs in maternal and child health, we can also expect several other key questions to be answered in the near future. First and foremost, while no individual IE can answer this question, a planned meta-analysis of the IEs in the portfolio *may* shed light on whether RBF programs affect maternal and child mortality. A crucial area of further inquiry will be the effect, if any, of RBF programs on equity and access to high quality healthcare. More evidence will continue to shed light on the effects of RBF programs on health worker motivation in varying contexts and at varying levels of incentive payments as well as on the staff retention effects of performance-based incentives. The impacts of RBF programs on clinical quality will likely also become clearer as more innovative measures are applied both programmatically as well as in IEs. In addition, several smaller parts of the RBF “black box” may also be unpacked using a combination of IE and administrative data. Such questions might include: what are the benefits of better record-keeping by providers, and better data reporting and greater autonomy at the facility level? We will likely also learn how RBF programs compare to other interventions, such as supportive supervision, continuous quality improvement, and enhanced financing. Finally, cost-effectiveness studies will help us understand the relative value of RBF programs compared to alternative interventions. Therefore, in the coming years, we expect the evidence from the HRITF IE portfolio to help build the evidence base for decision-making that leads to the universal access to high quality maternal and child healthcare.