

Impact Evaluation of Zambia's Health Results-Based Financing Pilot Project



Jed Friedman

Jumana Qamruddin

Collins Chansa

Ashis Kumar Das



@ 2016 World Bank Group
1818 H Street NW
Washington, DC 20433
USA
All rights reserved.

This volume is a product of the staff of the World Bank Group. The findings, interpretations, and conclusions expressed in this paper do not necessarily reflect the views of the Executive Directors of the World Bank Group or the governments they represent. The World Bank Group does not guarantee the accuracy of the data included in this work. The boundaries, colours, denominations, and other information shown on any map in this work do not imply any judgment on the part of the World Bank Group concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

The material in this publication is copyrighted. Copying and/or transmitting portions or all of this work without permission may be a violation of applicable law. The World Bank Group encourages dissemination of its work and will normally grant permission to reproduce portions of the work promptly.

For permission to photocopy or reprint any part of this work, please send a request with complete information to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA, telephone 978-750-8400, fax 978-750-4470, <http://www.copyright.com/>.

All other queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA, fax 202-522-2422, e-mail pubrights@worldbank.org

Table of Contents

LIST OF ABBREVIATIONS AND ACRONYMS	V
ACKNOWLEDGMENTS	2
EXECUTIVE SUMMARY	3
1. BACKGROUND AND PROGRAMME DESIGN	9
1.1 ZAMBIA HEALTH RESULTS BASED FINANCING (HRBF) PROJECT OBJECTIVES	10
1.2 ZAMBIA RBF MODEL AND DESIGN	11
2. IMPACT EVALUATION DESIGN	14
2.1 SELECTION OF DISTRICTS FOR THE IMPACT EVALUATION	15
2.2 RESEARCH QUESTIONS	16
2.3 METHODS AND DATA – IMPACT EVALUATION	16
2.3.1 Health Facility Survey.....	17
2.3.2 Household Survey	18
2.3.3 Estimation of standard errors for highly clustered population level outcomes	18
3. COUNTER EXTERNAL VERIFICATION	19
4. COST-EFFECTIVENESS ANALYSIS	20
4.1 MEASUREMENT OF COSTS	20
4.2 MEASUREMENT OF EFFECTIVENESS	20
4.3 COST-EFFECTIVENESS ANALYSIS	22
5. RESULTS	23
5.1 HEALTHCARE COVERAGE	23
5.2 QUALITY OF SERVICES.....	26
5.2.1 Structural Quality.....	26
5.2.2 Process Quality	30
5.3 EFFECT OF RBF AND ENHANCED FINANCING ON HEALTH SYSTEM PERFORMANCE MEASURES	34
5.3.1 Level of Revenue in RBF Health Facilities, GRZ vs RBF Grants.....	34
5.3.2 Level of RBF revenue, RBF vs C1.....	35
5.3.3 Allocation and use of RBF funds, RBF vs C1	36
5.3.4 Governance and managerial autonomy at health facilities.....	38
5.3.5 Satisfaction and motivation of the health workers.....	40
5.4 COST-EFFECTIVENESS ANALYSIS	41
6. DISCUSSION AND CONCLUSION.....	47
6.1 LESSONS LEARNT	53
7. LIMITATIONS OF THE STUDY	55
REFERENCES.....	56
APPENDIX 1	58
RESEARCH QUESTION 2 – DO HIGHER INCENTIVE PAYMENTS IN REMOTE AREAS RESULT IN INCREASED HEALTH OUTCOMES AND GREATER RETENTION OF STAFF?	58
A1.1 QUALITY OF SERVICES	58
A1.1.1 Structural quality.....	58
A1.1.2 Process quality.....	61

<i>A1.1.3 Client satisfaction</i>	62
A1.2 EFFECT ON THE HEALTH SYSTEM	63
<i>A1.2.1 Facility governance and autonomy</i>	63
<i>A1.2.2 Satisfaction and motivation of the health workers</i>	65
A1.3 CONCLUSIONS AND DISCUSSIONS.....	66
APPENDIX 2	68
RESEARCH QUESTION 3 – HOW DOES THE LIKELIHOOD OF AUDIT/EXTERNAL VERIFICATION OF RESULTS AFFECT THE ACCURACY OF REPORTED DATA?	68
B1.1 DATA REQUIREMENTS, SOURCES, AND SELECTION OF HEALTH FACILITIES	68
B1.2 RESULTS	69
B1.3 CONCLUSIONS AND DISCUSSION	71
APPENDIX 3	72
TABLE C1. MAPPING OF FACILITY QUESTIONNAIRE INSTRUMENT INTO THE BALANCED SCORE CARD	72
APPENDIX 4	73
POPULATION OUTCOMES AND FISHER-EXACT STANDARD ERRORS	73

List of Figures

Figure 1: Zambia Health Results Based Financing Model	14
Figure 2: Distribution of Intervention and Control Districts by Province, Zambia.....	15
Figure 3. Relative importance of quality components for generating the quality index for each service	21
Figure 4: Illustration of functions for assessing impact of quality of care on health.....	22
Figure 5: Proportion of GRZ grant to RBF grant.....	35
Figure 6: Funds disbursed to C1 districts as compared to RBF districts	36
Figure 7: Use of RBF Funds, and Proportion of RBF staff incentives to Government staff salaries.....	37
Figure 8: Distribution of programme costs: RBF and C1 groups combined.....	42
Figure 9: Incremental costs per capita over 2.25 years among three groups (US\$).....	43

List of Tables

Table 1: Incentivized Indicators and Unit Prices	12
Table 2: Areas for Quality Assessment and Associated Weights	13
Table 3: RBF Intervention and Control Districts by Province and total Population	16
Table 4: In-Facility delivery indicators	24
Table 5: Antenatal care coverage	24
Table 6: Postnatal care coverage	24
Table 7: Family planning indicators	25
Table 8: Immunization Coverage for children aged 12-23 months	25
Table 9: Health seeking behaviour for general illness, separately for under-5s and over-5s	26
Table 10: Status of infrastructure	27
Table 11: Availability of drugs	28
Table 12: Availability of medical equipment	29
Table 13: Mapping the quality checklist at health facilities	30
Table 14: Knowledge of maternal health danger signs: Results from the household survey	30
Table 15: Process quality of antenatal care provided: Results from the household survey	31
Table 16: Process quality of antenatal care: Results from patient exit interviews	32
Table 17: Process quality of postnatal care provided: Results from the household survey	32
Table 18: Process quality of child health care: Results from exit interviews	33
Table 19: Client satisfaction in antenatal care: Results from exit interviews	33
Table 20: Client satisfaction on child health care: Results from exit interviews	34
Table 21: Community participation, supervision, and performance assessment at health facility level	39
Table 22: Managerial autonomy at health facility level	40
Table 23: Job satisfaction	41
Table 24: Motivation for work	41
Table 25: Consumables expenditures from MSL during the pre- and post-RBF periods in three groups (US\$)	42
Table 26: Coverage and quality of key maternal and child health services at baseline and endline ..	44
Table 27: Lives saved from the RBF programme in comparison with lower and upper bounds	45
Table 28: QALYs gained from the RBF programme in comparison with controls, with lower and upper bounds	45
Table 29: Incremental cost effectiveness ratios, with lower and upper bounds	46
Table D1: In-Facility delivery indicators	74
Table D2: Antenatal care coverage	74
Table D3: Postnatal care coverage	74
Table D4: Family planning indicators	74
Table D5: Immunization Coverage for children aged 12-23 months	74
Table D6: Health seeking behaviour for general illness, separately for under-5s and over-5s	74
Table D7: Knowledge of maternal health danger signs: Results from the household survey	75
Table D8: Process quality of antenatal care provided: Results from the household survey	75
Table D9: Process quality of postnatal care provided: Results from the household survey	75

List of Abbreviations and Acronyms

ACT	-	Artemisinin-based Combination Therapy
ANC	-	Antenatal care
AZT	-	Zidovudine
BCG	-	Bacillus Calmette-Guérin
BP	-	Blood Pressure
C1	-	Control 1
C2	-	Control 2
CEA	-	Cost Effectiveness Analysis
CHWs	-	Community Health Workers
CSAs	-	Census Supervisory Areas
CSO	-	Central Statistics Office
DHMT	-	District Health Management Team
DIDs	-	Difference in Differences
DMO	-	District Medical Office
DPT	-	Diphtheria, Pertussis and Tetanus
D-RBFSC	-	District RBF Steering Committee
EMOC	-	Emergency Obstetric Care
EmONC	-	Emergency Obstetric and Neonatal Care
FP	-	Family Planning
GNI	-	Gross National Income
GRZ	-	Government of the Republic of Zambia
HMIS	-	Health Management Information System
HRBF	-	Health Results Based Financing
HRITF	-	Health Results Innovation Trust Fund
ICER	-	Incremental Cost Effectiveness Analysis
IE	-	Impact Evaluation
IPT	-	Intermittent Preventive Treatment
LCMS	-	Living Conditions Monitoring Survey
LIST	-	Lives Saved Tool
LMIC	-	Low Middle Income Country
MCH	-	Mother and Child Health
MNCH	-	Maternal, Newborn and Child Health
MDG	-	Millennium Development Goals
MMR	-	Maternal Mortality Ratio
MOH	-	Ministry of Health
MSL	-	Medical Stores Limited
ORS	-	Oral Rehydration Salts
PE	-	Process Evaluation
PMU	-	Project Management Unit
PNC	-	Postnatal Care
P-RBFSC	-	Provincial RBF Steering Committees
PSU	-	Primary Sampling Unit
QALYs	-	Quality Adjusted Life Years
RBF	-	Results Based Financing
SEAs	-	Standard Enumeration Areas
SSUs	-	Secondary Sampling Unit
TBAs	-	Traditional Birth Attendants
U5MR	-	Under-five Mortality Rate
UNICEF	-	United Nations Children's Fund

Acknowledgments

This report was prepared jointly by the Development Research Group; and the Health, Nutrition, and Population (HNP) Global Practice of the World Bank Group. The principal authors are Jed Friedman, Jumana N. Qamruddin, Collins Chansa, and Ashis Kumar Das. Contributions on the cost-effectiveness and human resources for health analyses were also received from Ha Thi Hong Nguyen (HNP, World Bank); Wu Zeng and Donald S. Shepard (Brandeis University); Gordon Shen (City University of New York); and Nkenda Sachingongu (University of Zambia). Sean W. Dalby (HNP, World Bank); Chitalu Miriam Chama Chiliba and Abson Chompolola (University of Zambia); Gelson Tembo (Palm Associates Limited); Ceaser Cheelo and Gibson Masumbu (Zambia Institute for Policy Analysis and Research); and Jos Dusseljee (ETC) assisted with the data collection and analysis at various stages of the study. The peer reviewers for the report were: Damien de Walque and Eeshani Kandpal from the Development Research Group; and Paul Jacob Robyn and Ayodeji Oluwole Odutolu from the HNP Global Practice.

Lastly, the authors acknowledge the support of senior management and technical team of the Ministry of Health, Zambia; senior management of the HNP Global Practice, World Bank: Olusoji O. Adeyi and Magnus Lindelow; senior team and secretariat for the Global Financing Facility/Health Results Innovation Trust Fund, World Bank: Monique Vledder, Rama Lakshminarayanan, Dinesh M. Nair, and Petronella Vergeer; and members of the World Bank Zambia HNP team: Netsanet Walelign Workie, Musonda Rosemary Sunkutu, John Bosco Makumba, Gertrude Mulenga Banda, and Yvette M. Atkins.

Executive Summary

In low and middle-income countries, there are various interventions that can be used to improve health system functionality and priority health outcomes. Results Based Financing (RBF) is one approach increasingly utilized in various countries across different settings and levels to facilitate these improvements.

This report reviews a comprehensive impact evaluation of a RBF pilot project in Zambia. The main objective of the paper is to present and analyse the IE results. While this report touches on some of the broader policy implications of this work, a separate brief has been developed detailing the policy implications to inform strategy and operations in Zambia and other countries. And while there have been a number of impact evaluations studies over the past few years measuring the effectiveness of RBF programs, the Zambia study is one of the few with a three-arm evaluation that tests RBF against an enhanced financing arm and a pure control. As such, this study provides some new and interesting insights on what can be effective in improving health systems and health outcomes.

Background and Programme Design

In an attempt to strengthen the health system and improve health-service delivery, Zambia has been gradually introducing RBF approaches to complement traditional input based financing in some of its health programs and activities. Zambia was awarded a US\$17 million grant in 2008 by the World Bank through the Health Results Innovation Trust Fund (HRITF) to implement a RBF pilot project with an accompanying Impact Evaluation (IE) led by the World Bank. Motivated by inadequate progress to achieving MDGs 4 and 5 targets, the primary objective of the project was to catalyse the country's efforts to reduce under-five and maternal mortality in 11 districts in nine (9) of Zambia's 10 provinces (except Lusaka) countrywide.

The Zambia health RBF (HRBF) pilot project was implemented by the Government through the Zambian health system (contracted-in) and is one of the few examples of a Lower Middle Income Country (LMIC) with this type of model. After a pre-pilot phase, which lasted approximately 2 years in the Eastern Province district of Katete, the RBF model was expanded to ten (10) additional districts in April 2012. By the end of the project, 203 health centres were covered across the country. This represented a total catchment population of about 1.5 million people of which the direct beneficiaries were 338,248 children aged between 0-59 months, and 372,073 women of child-bearing age.

The accompanying IE comprised both quantitative and qualitative approaches. Quantitative data for the IE at household and facility level was collected at baseline, implementation stage, and endline from 10 RBF intervention districts; 10 Control 1 (C1) districts; and 10 Control 2 (C2) districts. The method of selecting districts for the IE was based on district-matched randomization. Inputs were assigned to the three district groups as follows: (a) The RBF Intervention group to receive Emergency Obstetric and Neonatal Care (EmONC) equipment and RBF performance-based grants; (b) The C1 group ("enhanced financing" arm) to receive EmONC equipment exactly as in the RBF and the equivalent in money of the average RBF performance-related grants as input financing; and (c) The

C2 (“pure control” arm) group to receive nothing. The IE was designed to evaluate the impact of the pilot to inform three main policy questions:

- 1) What is the causal effect of the Zambian HRBF on the population health indicators of interest?
- 2) Do higher incentive payments in remote areas result in increased health outcomes and greater retention of staff?
- 3) How does the likelihood of audit/external verification of results affect the accuracy of reported data?

The IE investigated the impact of the RBF over a broad range of targeted and non-targeted indicators related to maternal and child health services. Baseline household data was collected over the period November to December 2011. Endline data was collected between November 2014 and January 2015, using the same survey tools and in the same study areas and was undertaken in 18 IE districts (all of the study districts in six of the matched district triplets yielding information from 6 RBF districts, 6 C1 districts, and 6 C2 districts). For the health facility survey, baseline data was collected between October and November, 2011 and endline data was collected between November 2014 and January 2015. At baseline, 176 health facilities were surveyed whereas 210 health facilities were surveyed at endline.

The household and health facility surveys provide information on the first two research questions. A special study known as “counter external verification” was carried out to explore the third research question. In addition, the IE was complemented by a Process Evaluation aimed at generating valuable information on a continuous basis during the implementation period. Furthermore, a cost-effectiveness analysis was conducted to gather information on the relative costs and effectiveness of the RBF programme from a health system perspective.

Summary of the Results

Of the nine indicators directly targeted by the RBF programme through the incentive structure, seven were directly measured or proxied in the population¹. Some of the measures responded to the RBF programme, with a broadly similar set also showing improvements under the enhanced financing arm (C1). Institutional deliveries in RBF districts increased by approximately 13 percentage points relative to the pure control districts (C2). However, the same indicator rose by about 18 percentage points in C1 districts relative to C2. Results for deliveries by skilled providers also show improvements in both the RBF and C1 districts relative to C2, but the C1 arm had greater magnitude for this indicator. One of the most important gains in the RBF arm was the timing of the first ANC visit which was earlier by two weeks as compared to both controls. This is an important gain in maternal care that is seldom observed in a broad-based primary care intervention such as RBF. On the other hand, the rate of change for any PNC was more rapid in C1 districts as compared to RBF districts. For immunization, full vaccination coverage declined in both C1 and C2 districts but remained constant or slightly higher in RBF districts, suggesting that the RBF programme could have

¹One of the remaining indicators involved special sub-population of HIV positive pregnant women whose coverage was not tracked by the data collection. The other remaining indicator applies to all women of reproductive age and not just those

been protective with respect to some measures of immunization coverage. However, the relative protective effects are not precisely estimated.

For structural quality, the quality of delivery rooms in RBF facilities were better than the delivery rooms in C1 and C2 districts. On process quality, the results were varied. For example, women residing in RBF districts were significantly more likely to list several out of the 12 danger signs during pregnancy as compared to those residing in C1 districts who were not able to list any. However, C1 districts witnessed better improvements in blood tests and any iron taken during ANC as well as higher immediate initiation of breastfeeding and receipt of Vitamin A after delivery as compared to both RBF and C2. On perceived quality, health workers in RBF facilities spent significantly more time during consultations with their patients as compared to both C1 and C2 health facilities.

On health systems, results suggest that RBF districts performed better than C1 districts in most of the indicators specifically availability of equipment, structural quality, managerial autonomy, accuracy in reporting, satisfaction and retention of health workers, and level and predictability of funding. A key salient point from the design perspective is that disbursement of funds directly to health facilities in RBF districts facilitated fiscal decentralization but the proportion of RBF funds which were disbursed to C1 districts as compared to the RBF districts was only 56%.

Results from the cost-effectiveness analysis (CEA) shows that the RBF (vs C1) provided more total health benefits (QALYs gained) but at a higher unit price. Nonetheless, in comparison with the two control groups, the RBF programme is a cost-effective approach in improving maternal and child health. When the RBF group is compared with the C1 group, the mid-point ICER is \$1,642 per QALYs gained (without quality adjustment), and \$1,324 per QALYs gained (with quality adjustment). When the RBF group is compared with the C2 group, the mid-point ICER is \$999 per QALYs gained (without quality adjustment), and \$809 per QALYs gained (with quality adjustment). All these values are less than the GDP/capita of \$1,759 in 2013 (mid-year of RBF programme) in Zambia^{2,3}. Since these ICERs are less than Zambia's GDP per capita in 2013, the RBF programme was cost-effective in comparison to C1 and C2. However, the input financing approach (C1) was also cost-effective in comparison to C2. The ICERs for C1 vs C2 were \$508 and \$413 per QALY gained, without and with quality adjustment, respectively. Thus, depending on which group is used for comparison, the ICER varies, but the estimates point to a cost-effective impact of both the RBF and C1 groups than the C2 group.

Conclusion

The overarching conclusion is that both the RBF and C1 interventions contributed to increased utilization of key MNCH services in Zambia. However, as compared to the C1, RBF had a more positive effect on health systems functionality and governance. Internal and external verification of results, and regular support supervision which were a key feature in the RBF districts was also helpful in tracking the performance of health workers and the programme as a whole. Another feature is that purchasing mechanisms were enhanced in the RBF and this potentially contributed to

² The World Bank. GDP per capita (current US\$). Washington, DC: The World Bank; 2015 [cited 2015 Sept 30]; Available from: <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>.

³ WHO recommends comparing ICER to GDP/Capita. GDP/Capita proxies for the productivity a person in a year. If an intervention could save more than what a person produces in a year, it is regarded as highly cost-effective.

greater efficiency and value for money. From the CEA analysis, we can conclude that the RBF (vs C1) provided more total health benefits (QALYs gained) but at a higher unit price. However, both the RBF and C1 are cost-effective at improving maternal and child health as compared to C2.

In the discussion section of the main report, we examined some of the causal and behavioural mechanisms through which the RBF and C1 could have achieved and/or not achieved gains in the targeted indicators. For the enhanced financing (C1) arm, the key question is whether the gains were as a result of increased financing, earmarking of funds for priority maternal and child health interventions, or other factors. A corollary question is whether greater gains could have been observed in the C1 arm if financial flows to C1 facilities actually equalled those received by RBF facilities. We observe that health facilities in C1 districts may also have been implementing RBF initiatives. On the other hand, with no concealed investigation, the study units were aware of the experiment and the C1 districts could have tried to out-perform the RBF districts. For the RBF, the key question is whether the RBF districts could have achieved much more. In exploring this question, we noted that the Zambia RBF project was being implemented in a health system that already had high coverage in some of the key MNCH indicators. As such, perhaps it would have been more prudent to have adopted a target or coverage based performance incentive framework rather than fee-for-service. Furthermore, the results show that health workers received about 10% of their official GRZ staff salaries on average as RBF staff incentives. This may not have been sufficient to have induced change.

Lessons Learnt

The results from the study shed light on several grey areas which have been under discussion with regards to the RBF in terms of the results, implementation and evaluation processes:

- (i) The study demonstrates that an RBF programme can be successfully implemented to increase delivery of key health indicators through “contracting-in” a capacity constrained public health system. Many other examples of successful public sector RBF programs occur in middle-income countries (i.e. Argentina) or when implemented by a specialist third party (i.e. Zimbabwe). Since Zambia implemented the RBF by using existing government systems, structures, and local expertise, it is potentially easier to scale-up a countrywide RBF programme. This is because the Zambia RBF design allows for financial, institutional, and impact sustainability.
- (ii) It is important to have a routine process evaluation (PE) system in place to continuously monitor the results and overall implementation of the RBF programme. The Zambia RBF pilot programme benefited from a PE system which provided regular updates and insights on the implementation to allow for midcourse corrections and evidence-based policy and planning.
- (iii) While the “contracting-in” design could be potentially more institutionally sustainable, consistency in leadership is a critical component to moving from a project to a programmatic approach that is fully embedded in the larger health sector. In the case of Zambia, there were several exogenous shocks (governance issues, split of ministries, high staff turnover at all levels of Government etc) which made it difficult to have continuous policy dialogue on RBF. To help ensure integration of experiences and lessons of current

and future RBF programs into broader health sector dialogue, these programs should be firmly embedded in the planning department of the Ministry of Health (MOH) with co-leadership with a relevant technical unit such as Mother and Child Health. The implementation structure could consist of a mix of dedicated technical civil service staff. In addition, an RBF coordination committee governed by the MOH should bring together interested donors together with Government to discuss emerging results, policy impacts, and the way forward.

- (iv) The Zambia RBF project demonstrates that having a performance incentive framework (provider payment mechanism) linked to targets and production capacity instead of a payment mechanism for all services rendered is potentially better. The Zambia RBF was implemented in a health system that already had relatively high coverage in some of the key maternal and child health indicators. As such, rather than fee-for-service, it may have been more effective to have used a target or coverage based performance incentive framework.
- (v) The enhanced financing arm as part of the evaluation is critical in order to be certain that effects in the RBF arm are not only due to additional financial resources. As shown in the evaluation results, enhanced financing, can also produce good results. In Zambia, these results go a step further in highlighting a potential issue in the current health system related to funding constraints and unpredictability. Input financing with parameters focused on key interventions can work—but in the case of Zambia, there were also issues in utilizing funds in Control 1 districts which points to disbursement mechanism issues which were not experienced when disbursing RBF grants in the RBF arm. Thus, effective approaches, including direct disbursement of funds to front-line service delivery levels coupled with a variety of financing mechanisms can have a positive impact on service delivery through improved budget performance (disbursement and absorption of funds).
- (vi) Direct disbursement of funds to front-line service delivery levels and use of an effective disbursement mechanism can also increase predictability of funding and managerial autonomy. However, the RBF funds may have substituted rather than complemented government funding due to the poor disbursement of Government grants to pilot health facilities by the district management during the implementation of the RBF project. To mitigate this problem, future RBF programs in Zambia could consider putting in place indicators linked to government budget performance at national, provincial, and district levels to ensure that RBF grants at service delivery levels are additional to government grants.
- (vii) Adequate levels of incentives need be offered to providers to trigger sufficient behavioural change. The relatively low power of RBF staff incentives in relation to guaranteed individual staff salaries (which declined to 10% over the project period) may have limited some of the possible achievable gains by RBF. Furthermore, higher incentive payments in remote areas did not result in increased health outcomes either (Appendix 1). This suggests that provider effort may be relatively inelastic at small incentive levels. This probably explains why the RBF had no impact on the motivation of health workers but had a positive impact on health worker satisfaction, reduced attrition, and responsiveness to the client. Given the high cost of living in Zambia, the additional income from the RBF staff

incentive could have been inadequate to fully influence personal behaviour. Future RBF programs should provide adequate but sustainable levels of RBF staff incentives.

- (viii) When introducing the concept of data verification in a health system with little previous experience, repeated outreach to facility management combined with experiential learning may be necessary for management to internalize the reality of a verification audit. This also applies to the possible ramifications for mis-reporting. The audit experiment discussed in Appendix 2 demonstrates a very low level of understanding of the audit likelihoods despite repeated announcement to the facility management, as well as a lack of understanding over mis-reporting thresholds and possible sanctions. As such, the audit likelihood experiment largely failed as the reporting principals were unaware of the likelihood assigned to the facility. Nevertheless, despite discrepancies in reporting found in RBF facilities by the external verifiers, these discrepancies appear to be within the bounds of normal reporting error as they are not significantly different from a sample of C2 facilities.
- (ix) A key component of the Zambia HRBF IE is the cost-effectiveness analysis (CEA) which justified the value of the RBF on both the costs and effectiveness (by increasing both quality and quantity of services). By adding a complementary cost-effectiveness study, the Zambia HRBF IE showed that a number of decisions must be made in the health facilities on health systems inputs such as personnel, drugs, equipment, buildings, verification, supportive supervision and so forth. The existence of both fixed and variable costs are important aspects in evaluating how much it costs to implement a RBF programme, and the efficacy of RBF programs as compared to non-RBF programs.
- (x) To our knowledge, this CEA study is among the few to incorporate the quality of care in the cost-effectiveness modelling, and the study innovatively uses a Delphi panel to generate a quality index from household survey based results and to convert a quality index to a health effect index. Given that improving the quality of care is one of the major components of the RBF programme, RBF evaluation models should always include an assessment of quality improvements to fully estimate the cost-effectiveness of the RBF programme.

1. Background and Programme Design

1. Zambia is a lower-middle-income country with a GNI per capita of US\$ 1,760 (World Bank, Atlas method) in 2014. Total population is estimated at 15.02 million in 2014, 60% of which resides in the rural areas. Against a backdrop of consistent market-led economic policies, Zambia began recording consistently high economic growth of above 6% in 2006 which went to 7.3% in 2012 after which it reduced to 6.7% in 2013, and 5.6% in 2014. Notwithstanding the positive economic developments, poverty has persisted and income inequality is still high. The effect of economic growth on overall poverty reduction has been insignificant and the urban centred growth has not generated higher incomes and better basic services for rural residents. According to the Living Conditions Monitoring Survey (LCMS) of 2010, poverty levels remain very high with 60.5% of the population living below the poverty line and 42.3% in extreme poverty. While poverty in urban areas has declined from 29.7% in 2006 to 27.5% in 2010, it is still very high in rural areas (77.9% in 2010 compared to 80.3% in 2006). The rural poverty rate at 77.9% is more than double the urban poverty of rate 27.5% and, over the past decade, the Gini coefficient increased from 0.47 to 0.52 which reflects a high level of income inequality. Life expectancy at birth is 51.2 years, which is lower than the average for its income group (65.8 years) and the sub-Saharan Africa average (54.7 years).
2. In the health sector, Zambia has made notable progress in improving health and nutrition outcomes in the last decade. However, progress has been insufficient to achieve some of the health and nutrition related MDGs. While under-five mortality (U5MR) decreased from 119 to 75 per 1,000 live births between 2007 and 2013/14 (CSO *et al.* 2014), this is still high compared to the average for lower middle income countries (61 deaths per 1,000 live births) and insufficient to achieve MDG 4. The maternal mortality ratio (MMR) also fell from 591 to 398 per 100,000 live births between 2007 and 2013/14 (CSO *et al.* 2014), but this is significantly above the average for Zambia's income group (260 per 100,000 live births). Chronic malnutrition in under-5 children decreased from 45.4% in 2007 to 40.1% in 2013/14 (CSO *et al.* 2014) but this is far below the MDG target of 23%.
3. The other challenge is low coverage and utilization of high impact maternal, child health and nutrition services. For example, whereas 96% of pregnant women received any antenatal care from a skilled provider in 2013/14, only 67% of these women deliver in a health facility, and only 64% are attended to by a skilled provider (CSO *et al.* 2014). Some of the underlying causes of this are: inadequate and poorly motivated health workers; an erratic supply of essential medicines and medical supplies; limited autonomy in decision-making at decentralized levels of the health system; and a weak monitoring and evaluation system. Critical among all these challenges is a human resources for health crisis which is evident from the limited availability and mix of skilled human resources, and which has contributed to an inequitable mix, absenteeism, tardiness, and poor morale among the health workers. The 2009 Health Public Expenditure Review observed high rates of absenteeism (21 percent self-reported) and tardiness (43 percent self-reported), equivalent to a total loss of 4,108 working days per month. Eliminating absenteeism and tardiness would translate to a gain of 187 full-time equivalent staff, enough to staff 21 rural health centres in Zambia.

4. In an attempt to strengthen the health system and improve health-service delivery, Zambia is gradually introducing Results-Based Financing (RBF) approaches in financing some of its health programs and activities to complement traditional input based financing. In RBF, *“a principal entity provides a financial or in-kind reward, conditional on the recipient achieving pre-agreed actions and performance goals.”* In principle, RBF can encompass various forms of output-based aid, provider or healthcare based incentives for performance and consumer incentives for behavioural change. By introducing incentives that reward results, RBF promotes greater accountability of service providers, and improves management, efficiency, equity of service delivery, and health information systems with the overall aim of strengthening service delivery to improve development outcomes.
5. RBF has been advocated as a key transformative approach to health financing with potential to strengthen health systems, and improve health outcomes. Existing evidence shows that RBF can help to strengthen health systems by decreasing costs of service provision, improving staff motivation and morale through the provision of staff incentives, and empowering providers and beneficiaries in the use of data for decision making, and decentralisation of health services. Countries that have experienced increases in service availability and, for some dimensions of service quality include Rwanda, Argentina, and Zimbabwe (Basinga et al. 2011; Gertler et al. 2014; Friedman et al. 2015). However not all attempts to experiment with RBF have resulted in increased population coverage and health system improvements. These examples include Afghanistan and DRC (Alonge et al. 2014; Huillery et al. 2014). In these latter cases, often difficulties in implementation appear to have contributed to the limited effectiveness.
6. At design stage, the Zambia Government was particularly interested in testing a “Contracting-in model to strengthen aspects of the public health system. This was motivated by a long history of Performance-Based Contracting (1996-2006) and the Government was determined to apply RBF within the public structures which were left behind when the Central Board of Health (the purchasing agency at that time) was abolished in 2006. Therefore, the Zambian HRBF project was implemented through the Zambian health system (contracted-in) and is one of the first examples using this type of model globally.

1.1 Zambia Health Results Based Financing (HRBF) Project Objectives

7. Zambia was awarded a US\$17 million grant in 2008 by the World Bank through the Health Results Innovation Trust Fund (HRITF)⁴ to implement a Results Based Financing (RBF) project as well as an Impact Evaluation. The prime objective of the Zambia HRBF project was to catalyse the country’s efforts to reduce under-five and maternal mortality in 11 districts in nine (9) provinces (except Lusaka) countrywide. This project was motivated by the slow progress in MDGs 4 and 5 targets, and particularly:

⁴ At the time of the Zambia HRBF project, the HRITF was being funded by the British and Norwegian Governments.

- i. Critical human resource shortages both in terms of numbers and skills-mix, and low productivity;
 - ii. Poor quality of service provision and inequities by sex, age, income, education, type of service etc;
 - iii. Varied intervention coverage, and supply chain dysfunctions across interventions and geographical locations; and
 - iv. Poor governance and the need for increased value for money i.e. strengthening the link between financing and health outcomes.
8. The anticipation was that the Zambia HRBF project would contribute to the delivery of better quality services by increasing the productivity of the existing human resource base, augmenting financial resources, increasing managerial autonomy and decision-making of the implementing health facilities, and improving transparency and accountability in the use of funds. A pre-pilot project was implemented from 2009 to 2011 in Katete District in the Eastern Province of Zambia with the goal of developing an RBF operational design and model specific to the context and prevailing situation in Zambia. After the Katete pre-pilot, the RBF model was expanded to ten (10) additional districts in April 2012. By the end of the project, 203 health centres countrywide were covered. This represents a total catchment population of about 1.5 million people of which the direct beneficiaries were 338,248 children aged between 0-59 months, and 372,073 women of child-bearing age.

1.2 Zambia RBF Model and Design

9. The Zambian HRBF project was implemented through the Zambian health system (contracted-in) and is one of the first examples globally of this type of model. The overall activities which were supported by the project include:
- i. Provision of incentive payments to both district level management and health facilities based on individual and institutional performance;
 - ii. Provision of a package of reproductive health commodities and equipment;
 - iii. Supporting evidence-based decision making by improving the availability and quality of data generated through the Health Management Information System (HMIS). This includes data management, data analysis, reporting, and use;
 - iv. Enhancing quality in the provision of health care through regular and rigorous clinical quality assessment, and perceived quality through periodic client tracer surveys;
 - v. Capacity building and training in the delivery of Reproductive Health services - specifically Emergency Obstetric and Neonatal Care (EmONC); and
 - vi. Training in planning and budgeting, and financial and procurement management.
10. The Zambia HRBF model (Figure 1) adhered to the principle of “separation of functions” and performance was rewarded based on nine (9) facility based maternal and child health output indicators, and ten (10) dimensions of quality. Performance-based payments were provided to health facilities and districts after the attainment of the pre-agreed indicators on quantity and quality. The indicators were: i) institutional deliveries by a skilled birth attendant; ii) curative consultation, iii) ANC prenatal and follow up visits; iv) postnatal visit; v) full immunization of children under one year; vi) pregnant women receiving 3 doses of malaria IPT; vii) family planning users of modern methods at the end of the month; viii) pregnant women counselled

and tested for HIV; and ix) HIV pregnant women given Niverapine and AZT. This is shown in Table 1. The ten service areas targeted for quality improvements are comprised the following: Curative Care, Antenatal Care, Family Planning, Immunization, Delivery Room, HIV/AIDS, Supply Management, General Management, Health Management Information System, and Community Participation. This is shown in Table 2.

11. In addition to the Health Centres, the District Medical Office (DMO) was rewarded if it fulfilled a set of supervision and management functions. The incentives were tied to strengthening the DMO's role in supporting health facilities' efforts to increase the delivery of high-quality services based on their performance on management and supervisory functions. According to the stipulated roles and functions, Health Centres were the frontline service providers, while the District Medical Offices and District (General) Hospitals were the quantity and quality assessors, respectively. Through internal and external verification processes, reported data was extensively audited. District RBF Steering Committees (D-RBFSC) were the internal verifiers, Provincial RBF Steering Committees (P-RBFSC) were purchasers, and the Ministry of Health (MOH) headquarters was both the fund holder and regulator. The steering committees at provincial and district levels was an assembly of a wide range of stakeholders from other Government Ministries and Departments, Civil Society Organisations, NGOs, and the beneficiary communities.

Table 1: Incentivized Indicators and Unit Prices

Indicator	Unit Price (US\$)
1 Curative Consultation	0.2
2 Institutional Deliveries by Skilled Birth Attendant	6.4
3 Antenatal Care (prenatal and follow up visits)	1.6
4 Postnatal visit	3.3
5 Full immunization of children under one year	2.3
6 Pregnant women receiving 3 doses of malaria IPT	1.6
7 Family Planning users of modern methods at the end of the month	0.6
8 Pregnant women counselled and tested for HIV	1.8
9 Number of HIV pregnant women given anti-retroviral therapy prophylaxis (Niverapine and AZT)	2.0

Source: MOH (2011): Zambia HRFB Project Implementation Manual

Table 2: Areas for Quality Assessment and Associated Weights

Service Area	Weight
Curative Care	0.11
Antenatal Care	0.16
Family Planning	0.18
Expanded Programme on Immunization	0.09
Delivery Room	0.20
HIV Services	0.05
Supply Management	0.07
General Management	0.06
Health Management Information System	0.06
Community Participation	0.03
Total	1.00

Source: MOH (2011): *Zambia HRF Project Implementation Manual*

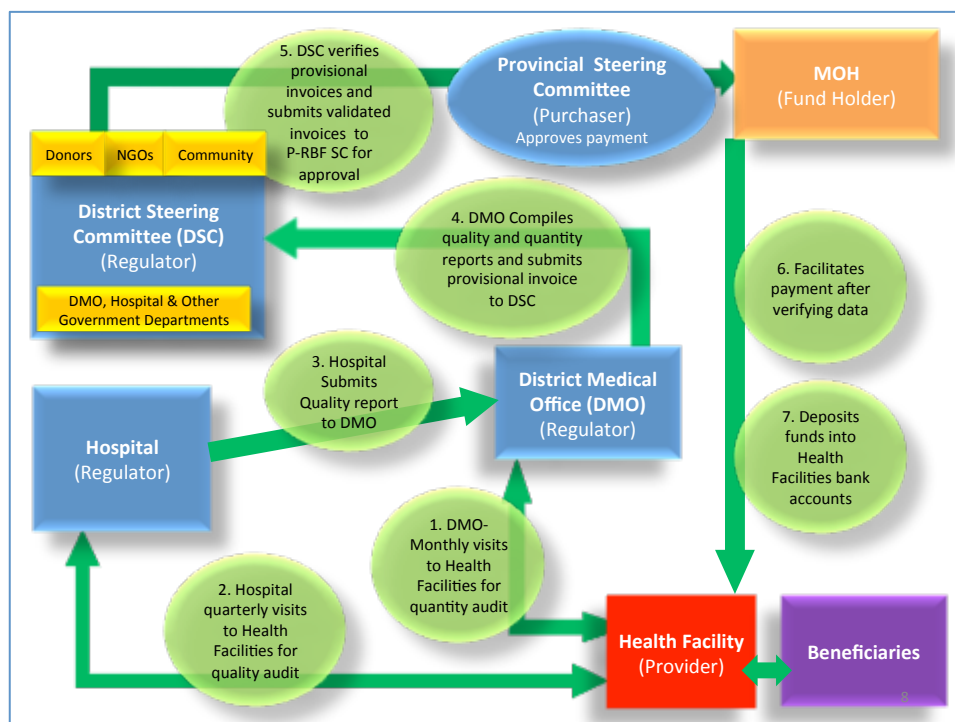
12. Two external verification exercises were conducted during the course of the project aimed at independently assessing the completeness, accuracy and validity of self-reported and internally verified data at the health facilities. Data on all the nine (9) incentivized health facility output indicators were verified. Key sources of data at the health facility level were the Health Information Aggregation (HIA) 2 forms in which the facility summarized (aggregated) the level of service provision per indicator. Since this data required verification, tally sheets, activity sheets and registers were used as these indicated the individual services delivered to each client. This was relevant in that it provided further details as to the date of the service, client register number, name of client, residential address, among others. Qualitative aspects of service delivery were also reviewed in order to put into context the level of service delivery. The external verification also included client tracer surveys where clients were selected from health facility records (registers) for tracing into the community and confirming that services were received as well as to measure perceived quality. The client tracer surveys focused on two indicators, namely: (i) Deliveries by skilled personnel, and (ii) full immunization of children below the age of one (1) year. Verbal responses on service-related questions were also obtained from clients during the tracer surveys.
13. To promote fiscal decentralisation and support autonomy of resources, health facilities on the RBF received performance-based payments directly into their bank accounts after the delivery of the pre-agreed indicators on quantity and quality⁵. The health facilities were authorised to use at least 40% of the money they earned from the RBF for operational activities (both on the supply and demand-side) to increase the number of services being delivered⁶. A maximum of 60% of the money could be used for staff motivation bonuses.

⁵ A Threshold-Based Graduated method was being used to calculate the final RBF payment due to the health facility. The health facility was rewarded for quality if the quality score was 61% and above as follows: (i) 61%-69% (Extra 15% of quantity payment); (ii) 70%-79% (Extra 25% of quantity payment); and (iii) 80%-100% (Extra 50% of quantity payment)

⁶ This could include purchase of safe delivery kits, upkeep of the health facility, community outreach, contracting of retired nurses and midwives, etc.

14. There was a strong focus on the quality of services provided as quarterly assessments were conducted to enhance quality of service provision. DMOs subcontracted hospitals to undertake these assessments based on a comprehensive quality checklist as outlined in Table 2 above. The quality assessment checklist incorporated international and national operating standard operating practices and guidelines in ten different areas deemed critical for quality improvement and assurance. The timing of the quarterly quality assessments were unannounced to the health facility teams and could take place anytime during a particular quarterly period.

Figure 1: Zambia Health Results Based Financing Model



Source: MOH (2011): Zambia HRBF Project Implementation Manual

2. Impact Evaluation Design

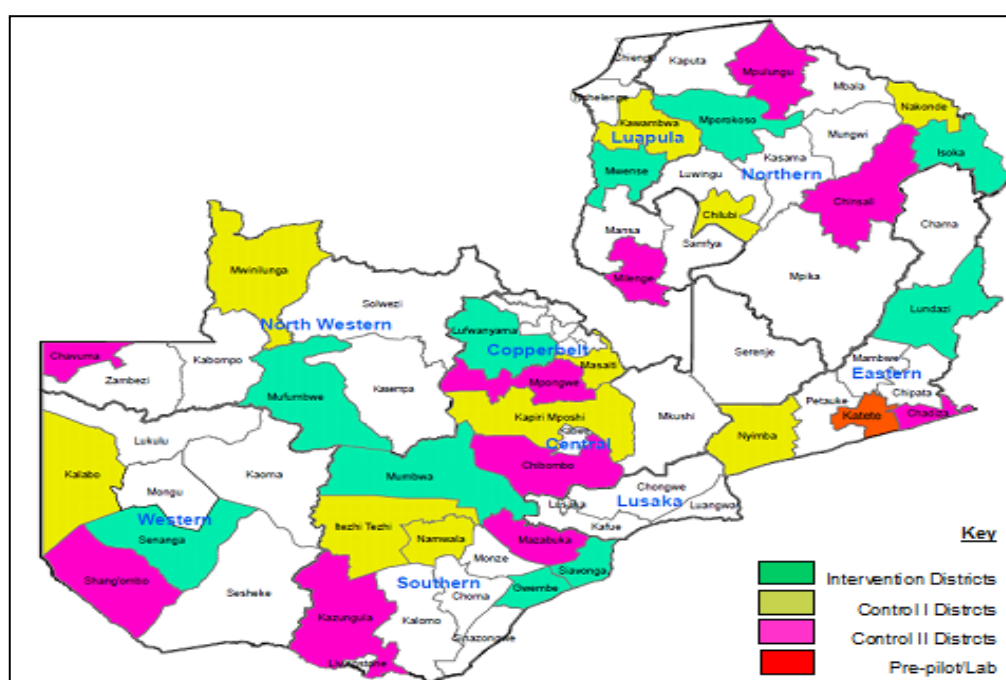
15. The Zambia HRBF project was accompanied by an Impact Evaluation (IE) component which comprised both quantitative and qualitative approaches. Dedicated primary quantitative data for the IE was collected at baseline, implementation stage, and at the end of the project from 10 RBF intervention districts; 10 Control 1 (C1) districts; and 10 Control 2 (C2) districts as shown in Figure 2 and Table 3. The method of selecting districts for the IE was based on district-matched randomization and is described in more detail below. For the Zambia IE, the districts were chosen as the unit of randomization and not individual health facilities because the DMO plays a key role in supervision/monitoring and is itself incentivized. Randomization at the facility level wouldn't have been successful because numerous channels and spill overs could have affected the IE design and implementation. Inputs were assigned to the three district groups as follows: (a) The RBF Intervention group received Emergency Obstetric and Neonatal Care (EmONC) equipment and RBF incentives; (b) The C1 group received EmONC equipment exactly as in the

RBF and the intended equivalent of the average RBF incentives in the RBF Intervention group; and (c) The C2 group received nothing.

2.1 Selection of districts for the Impact Evaluation

16. The evaluation team, in consultation with government, decided that districts selected for the HRBF study should approximate the median population health, socio-economic condition, and health governance capacity for the collection of districts in the provinces in which they are located. If the evaluation instead focused on exceptionally high (or exceptionally low) capacity districts, then this may in turn overstate (or understate) the estimated gains from a national scale-up for the RBF.
17. To select districts for the study, district level information was gathered on three areas of interest: district health administrative capacity, district population health service outcomes, and levels of district population living standards. The administrative capacity of the district was measured as an index derived from a principal components analysis based on the following three measures:
 - a) Average facility level stock-out rate of key commodities over the years 2006 and 2007
 - b) Average supervisory visit rate from District medical Office (DMO) to all facilities over the years 2006 and 2007
 - c) Rate of under-5 population covered by immunization campaigns in 2006 and 2007
18. Within each province (except for Northern and Southern Provinces, where six districts were sampled), three districts at or near the provincial median index score derived from these measures were selected and then randomly assigned to each of the three arms. Thus, there are a total of 30 districts distributed equally among the three study arms with 10 districts in each.

Figure 2: Distribution of Intervention and Control Districts by Province, Zambia



Source: MOH (2011): Zambia HRBF Project Implementation Manual

Table 3: RBF Intervention and Control Districts by Province and total Population⁷

Province	RBF Intervention	Population ⁸	Control 1 (C1)	Population	Control 2 (C2)	Population
Central	Mumbwa	226,171	Kapiri Mposhi	253,786	Chibombo	303,519
Copperbelt	Lufwanyama	78,503	Masaiti	103,857	Mpongwe	93,380
Eastern	Lundazi	323,870	Nyimba	85,025	Chadiza	107,327
Luapula	Mwense	119,841	Kawambwa	134,414	Milenge	43,337
Northern	Mporokoso	98,842	Chilubi	81,248	Chinsali	146,518
	Isoka	72,189	Nakonde	119,708	Mpulungu	98,073
North-Western	Mufumbwe	58,062	Mwinilunga	104,317	Chavuma	35,041
Southern	Siavonga	90,213	Namwala	102,866	Mazabuka	230,972
	Gwembe	53,117	Itezhi-tezhi	68,599	Kazungula	104,731
Western	Senanga	126,506	Kalabo	128,904	Shangombo	93,303
TOTAL		1,247,314		1,182,724		1,256,201

2.2 Research Questions

19. The IE was designed to inform policy on the impact of the project based on three key policy questions:

- What is the causal effect of the Zambian HRBF on the population health indicators of interest?
- Do higher incentive payments in remote areas result in increased health outcomes and greater retention of staff?
- How does the likelihood of audit/external verification of results affect the accuracy of reported data?

20. This overview report focuses on the first research question, while also addressing the second and third research questions through dedicated annexes.

2.3 Methods and Data – Impact Evaluation

21. Household and health facility surveys provide information on the first two research questions, whereas a special study known as “counter external verification” was carried out to explore the third research question. In addition, the IE was complemented by a Process Evaluation aimed at generating valuable information on a continuous basis during the implementation stage, and a cost-effectiveness analysis.

22. In terms of the empirical strategy adopted to estimate the causal impact of the intervention on the priority research questions, the first question is addressed through the quasi-experimental evaluation method of district-level matched difference-in-differences. To select these districts as well as comparison districts for study, three districts in each rural province (except for Northern and Southern provinces which provided six districts each) were selected according to a criterion

⁷ Based on the old administrative classification of districts and provinces

⁸ The RBF project was implemented in 11 districts (including Katete the pre-pilot district) with a total population of 1.5 million. Katete was not part of the Impact Evaluation (IE) which brings down the population for the IE to 1.2 million

mentioned in the next sub-section. Within these triplets of districts, each of the districts was then randomly assigned to either treatment or to one of two control statuses (equivalent financing or business-as-usual)⁹.

23. For the second research question, health facilities designated as remote in the RBF Intervention arm were randomized in two different groups. The remote designation is based on Zambia's civil service classification of districts into urban, rural, or remote for purposes of determining the level of hardship allowances for civil servants. This classification takes into account the different levels of development within a district by designating certain parts of a district as urban (mostly the district administration centre), others as rural (outside a 20KM and 15KM radius of the administrative centre in cities and municipalities, respectively), and the rest as remote (the periphery of a district). The total number of remote and non-remote health facilities in the RBF Intervention group was then identified. From the total number of remote health facilities, a smaller number of remote health facilities were randomly selected and assigned to be receiving additional incentives. In particular, the prices of all the quantity indicators for this small group of remote health facilities were pegged at 25% more than the normal prices of indicators in the non-remote health facilities (25% more than the prices in Table 1).

2.3.1 Health Facility Survey

24. Two rounds of facility surveys were undertaken for the same set of rural health centres and health posts, i.e. baseline (October-November, 2011) and endline (November 2014 - January 2015). The facility survey consisted of a facility checklist; a health worker instrument; exit interview of sick children and antenatal women. The training for the data collection of the enumerators included a classroom didactic part and a field visit. The instruments were pre-tested and modified appropriately before the actual survey:

Facility survey checklist

Health facilities were selected by a simple random sampling technique. In the baseline, 176 health facilities were surveyed whereas 210 were surveyed in the endline. The facility checklist was used to collect information on infrastructure, administration, availability of basic drugs and equipment, governance and autonomy.

Health Worker Interviews

Up to two health workers were selected for the interview at each health facility. The criterion for selection was provision of maternal and child health care on the day of the interview. Sample sizes were 326 and 355 for the baseline and endline surveys, respectively. The instrument included questions on remuneration, knowledge, job satisfaction and motivation.

Exit Interviews

A patient exit interview assessed patient satisfaction and quality of care received for patients exiting ante-natal care and child health consultations. For child consultation, the child's parent

⁹ The Zambia HRBF was implemented in 11 districts. However, Katete district being a pre-pilot district was excluded from the IE. At the start of the RBF implementation, the country had 9 provinces. Thus, for the IE, the old administrative arrangements were maintained.

or caretaker was interviewed. Up to six clients were selected per service through a systematic random sampling strategy (based on caseload for the same day of previous week for the facility). Sample sizes were as follows: Child health (baseline 1,059 and endline 1,266), and antenatal care (baseline 893 and endline 1254).

2.3.2 Household Survey

25. The first round of household data was undertaken during November-December 2011, whereas the second round took place around November 2014 - January 2015, using the same tools. The survey was undertaken in 18 IE districts (all of the study districts in six of the matched district triplets yielding information from 6 RBF districts, 6 C1 districts, and 6 C2 districts).
26. For statistical purposes, each district in Zambia is subdivided into Census Supervisory Areas (CSAs), which in turn nests Standard Enumeration Areas (SEAs). Thus, for data collection purposes, the SEA is the smallest geographical unit above the household and is the primary sampling unit (PSU). The SEAs were sampled from the catchment areas of selected health facilities. The sampling frame of SEAs in each treatment arm was arrived at by digitally overlaying SEA maps (obtained from the CSO) with health facility catchment area maps. After grouping the PSUs by stratum (treatment vs control), the sample was then selected in two stages: i) selection of PSUs in the first stage using probability proportional to size, and ii) selection of 10 eligible households, or secondary sampling units (SSUs), in the second stage using systematic random sampling. Prior to household selection, a full PSU listing of eligible households (households with a pregnancy related outcome, i.e. live birth, stillbirth, abortion and miscarriage within the two years prior to the survey) was undertaken by the survey team in each cluster. At baseline, 3,064 households in the relevant districts were surveyed at baseline, and 3,500 households at follow up.
27. The household survey included questions on coverage of ante-natal, post-partum and post-natal care, child health, delivery outcomes, family planning, and general health-seeking behaviour, mothers' knowledge of healthcare, as well as out-of-pocket expenditure.

2.3.3 Estimation of standard errors for highly clustered population level outcomes

28. The evaluation design for the Zambia RBF pilot involved pair-matched randomization at the district level. Randomization at the district level does come with potential inferential cost in the power of the analysis as the number of units of randomization is limited. In the case of the Zambia RBF, the RBF and C1 interventions were each piloted in 10 districts around the country. However due to budgetary limitations, population data was only able to be collected in six districts in each study arm. The main report estimates standard errors for impact estimates by clustering at the PSU level. However, as implied above, there may be unobserved influences at the district level that lead to district-level correlations in impacts that ideally should be accounted for. However, this presents two analytic difficulties.
29. The first difficulty is the fact of relatively few study units for the analysis. Besides the challenge to inferential power by the relatively few number of study units, traditional approaches to standard error estimates, notably the cluster-robust standard error, may be downward biased

and thus over-reject the null hypothesis of no treatment effect (Cameron et al., 2008). To counteract this potential bias, the precision of statistical tests can also be assessed through Randomization Inference (RI) which assumes all observed outcomes and covariates to be fixed and generates the reference distribution of test statistics by modelling the treatment assignment as the sole random variable in the data (Ernst, 2005). RI compares the actual point estimate observed in the evaluation against the distribution of all conceivable point estimates as determined through permutation methods – where the actual statistic falls in this distribution determines the exact p-value. This one-tailed hypothesis test is considered an exact test because it does not require a large-sample approximation as randomization itself is the basis for inference and permutation methods have exhausted all possible treatment assignments across districts. An exact test has the added benefit that it does not impose distributional assumptions that are often behind approximations of reference distributions in standard hypothesis testing.

30. For population level impacts, this report will present the exact p-values estimated through randomization inference (as well as compare them with the asymptotic p-values estimated with clustering at the PSU level in Appendix 4). The tables show that indeed the precision of the inference is not as great with the exact p-values. Many impacts, if they were estimated with incorrect asymptotic standard errors would indeed be found to be precise. This raises the question of the acceptable level of precision for impacts to inform policy when an evaluation does not have high power. Given the that the first pilot of RBF in Zambia faced various implementation challenges, population data was unable to be collected on a broad basis, and the international evidence base for RBF mechanisms comprises only a handful of countries, this report argues that policy makers should consider exact p-values larger than traditional cut-off levels, say on the order of 0.15, as sufficiently precise to inform future policy directions. This report will denote population impacts estimated with a p-value of 0.15 or less with an asterisk.

3. Counter external verification

31. To explore the effect of likelihood of audit/external verification of RBF results on the accuracy of reported data (third research question), a “counter external verification” was conducted. Health facilities in the RBF Intervention group were randomized into three groups and given letters notifying them a varying likelihood of performance audit at the start of the RBF project. The first group received a letter indicating a 100% likelihood of audit, the second group a 30% likelihood of audit, and the third group a 10% likelihood of audit. From each category of likelihood of audit, 35 facilities were randomly sampled. Data on all the nine (9) incentivised health facility output indicators were verified for the six-month period from July 01, 2013 to December 31, 2013. Key sources of data at the health facility level were the Health Information Aggregation (HIA) 2 forms, tally sheets, activity sheets and registers. These documents were used to check for errors relating to summing, recording and data entry.

4. Cost-effectiveness analysis

32. To complement the Impact Evaluation, a cost-effectiveness analysis was conducted aimed at informing the Zambian Government and other development partners on both the relative costs and effectiveness of the RBF programme. Given that the cost-effectiveness analysis of RBF is primarily designed to inform donors, development partners, and the Ministry of Health (MOH) about the future implementation or expansion of RBF, this cost-effectiveness analysis used a *health system perspective*. This perspective considers aspects that are the most relevant to decision makers particularly the costs of delivering services, but not travel expenses of consumers nor the economic cost of their time in accessing services.

4.1 Measurement of costs

33. The analysis used financial costs rather than economic costs, as financiers and implementers are most interested in returns from direct financial investments. The financial approach recognizes that many decisions must be made within a time-limited period. Within the relevant period, for government facilities, most personnel, equipment, and building expenditures are fixed costs, while many consumables are variable costs. Thus for the cost analysis, we included RBF programme costs, which include costs incurred by the World Bank for designing, implementing and monitoring the RBF programme and costs of consumables (drugs and supplies).

34. Programme costs, which were primarily for administration of the RBF programme and incentive payments, were obtained from the World Bank Zambia office. The programme costs allocated to both the RBF and C1 groups were based on the administrative records of the Project Management Unit (PMU) at the MOH. The World Bank's headquarters costs were allocated to the RBF and C1 groups in proportion to the programme costs. Costs of consumables such as drugs and supplies, were obtained from a data set compiled by Medical Stores Limited (MSL) from January 2011 through June 2014, with January 2011 to March 2012 (5 quarters) as pre-RBF period and April 2012 to June 2014 as post-RBF period (9 quarters). The cost of consumables were calculated as cost per population per quarter, and then a difference in differences (DIDs) approach was used to determine the incremental costs of consumables. All the costs were measured in US dollars during the project implementation period (April 2012 to October 2014).

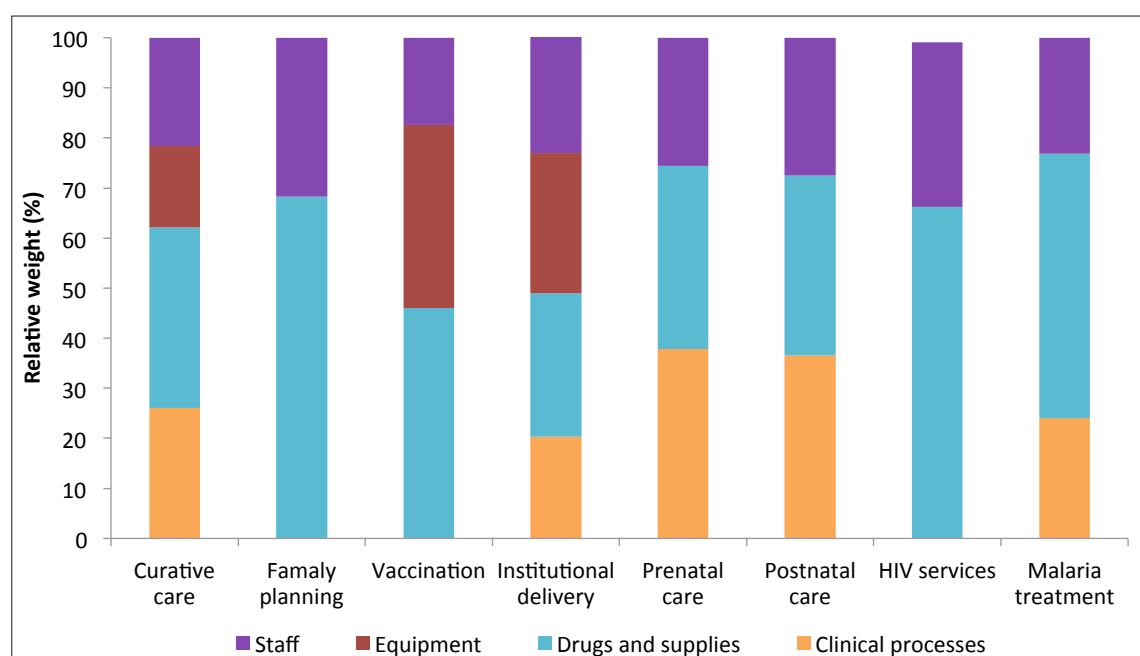
4.2 Measurement of effectiveness

35. To estimate effectiveness, the IE team obtained the results from both the household and health facility surveys from the impact evaluation, and selected those related to the RBF incentives that appeared to have improved as determined using a difference-in-difference (DID) approach. This included MCH services such as ANC, PNC, institutional delivery, IPT, family planning, and immunization. Out of all family planning methods, we observed that the use of injectable contraceptives had improved substantially, based on facility reported data, and hence was included in the analysis. We converted the utilization of injectable to the population coverage based on the population of reproductive age in the catchment area of health facilities and the expected use of injectable per year (4 times a year). The calculated national coverage of injectable was 21.9% in 2014, which is close to the estimate of 19.3% from the 2013/14 Zambia

Demographic and Health Survey (CSO et al, 2014). Most changes in curative services and HIV/AIDS services, as reported by the facility, were not statistically significant and thus were excluded from the analysis.

36. The Lives Saved Tool (LiST) was used to convert the coverage of health services to the number of lives saved (Avenir Health, 2015). The LiST tool has been widely applied in projecting the health impact of interventions and is advocated by the United Nations Children’s Fund (UNICEF) for decision making (Stenberg et al, 2014; Boschi-Pinto et al, 2010; Singh et al, 2014). To set up the LiST model for all control groups, the baseline coverage for control groups was adjusted to the same level as the RBF group, and the endline coverage of services was also adjusted accordingly to reflect the pure impact from DIDs. These baseline and endline coverages of selected services were used as inputs to the LiST to estimate the number of lives saved.
37. RBF is also expected to improve quality of the care. The quality of care was assessed from two rounds of the health facility survey, which measured general quality, clinical process, availability of drugs and suppliers, availability of equipment, and availability of qualified human resources. The IE team convened a Delphi panel with 11 experts with expertise of epidemiology and clinical medicine in Zambia in November 2014 to determine the relative importance of each quality component, and generated a quality index (ranging from 0 to 1) for each service. We conducted two rounds of a Delphi survey and used the results from the second round for the analysis. Figure 3 shows the relative importance of each quality component for different services.

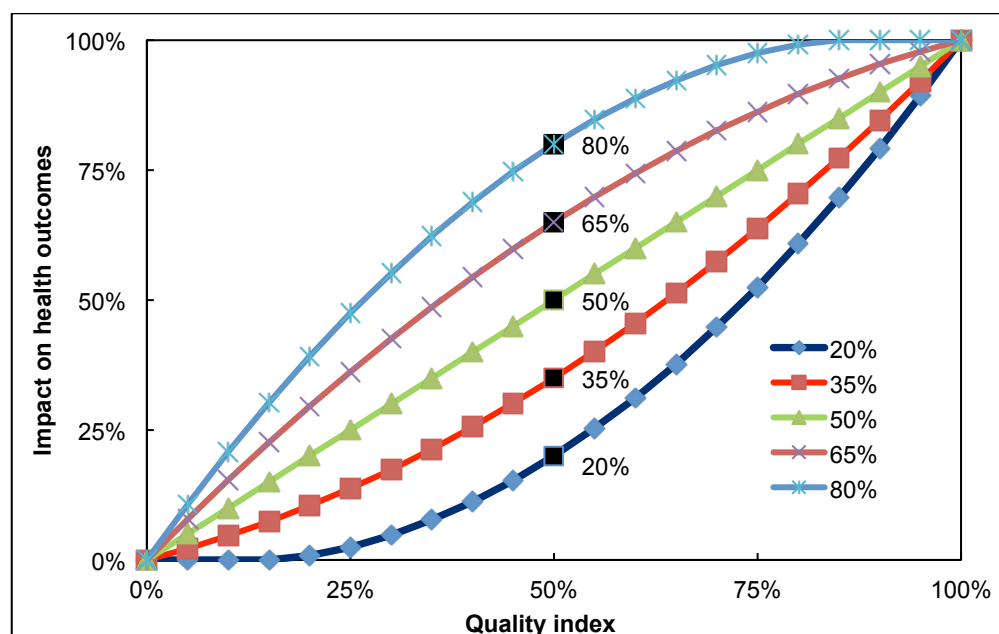
Figure 3. Relative importance of quality components for generating the quality index for each service



38. The IE team used the same expert panel to estimate the health impact of the quality of care to generate a health-effect index using a quadratic function (See Figure 4). The quadratic function was used because of its flexibility to accommodate concave up, concave down, and linear

relationships. To incorporate quality of care in the analysis, an effective coverage was generated by multiplying the health-effect index by the coverage of corresponding services from the household or health facility survey. The result was treated as quality adjusted coverage to feed into the LiST model.

Figure 4: Illustration of functions for assessing impact of quality of care on health



4.3 Cost-effectiveness analysis

39. The IE team used key parameters from the Zambia data preloaded in the LiST tool (e.g. the age structure of the population), and adjusted the population size to the size of the catchment population in the RBF group. The LiST tool produced the number of lives saved from improved intervention. We converted it into quality-adjusted life years (QALYs) employing the formula for fatal cases (Sassi, 2006) using Zambia's life table (WHO, 2015a; 2015b). Due to the difference of population sizes for costs, the IE team calculated incremental costs and effectiveness per capita, and generated the incremental cost-effectiveness ratio (ICER).

$$ICER = \frac{\text{Incremental costs of RBF program/capita}}{\text{QALYs gained/capita}}$$

40. A bound analysis was also conducted based on the lower and upper bound of the effectiveness of services in terms of improvement of the coverage of services. The lower bound estimate of the ICER used all lower bounds of the impact across the five services that are included in the analysis, while the upper bound estimate used all upper bounds of the same five services as used in lower bound estimates.

5. Results

41. This main section of this report provides the findings and conclusions on the first research question, while the results for questions 2 and 3 are presented the in Annex.

5.1 Healthcare coverage

42. This section evaluates the effect of the RBF intervention on key maternal and child health outcomes, using the results of the household survey with a focus on indicators targeted in the Fee-For-Service Package, and other related indicators. The “impact estimate” given is the difference in the change in the outcome for households in RBF and control districts between the baseline and endline household surveys, controlling for the stratification indicator of pair-matched districts. The listed p-value can be interpreted as the probability that the “true” impact – when the RBF intervention is compared to either C1 or C2 – is equal to zero. As explained earlier, the Fisher exact p-values are presented, and outcomes with a p-value less than 0.15 are indicated with an asterisk.

Institutional (In-Facility) deliveries and deliveries by skilled providers

43. Table 4 conveys the mean values of key delivery related outcomes at both baseline and endline survey periods for the three study areas: RBF, C1 (enhanced financing), and C2 (observational control). Alongside the mean values are the impact estimates for the RBF programme vis-à-vis each of the two control arms. These delivery outcomes, as well as the following tables that convey antenatal and postnatal care information, are based on individual recall of the mothers for all births within a two-year period before the date of interview. In other words, the baseline data seeks health information for the two years before programme onset, while the endline data seeks the same information for the two-year period of complete exposure to the RBF programme.
44. The results reveal that the in-facility delivery rate and the in-facility delivery rate by a skilled provider (with skilled providers defined as doctor, clinical officer, midwife, or nurse) increased between the baseline and endline in almost all sets of districts. For example, in RBF study districts the in-facility delivery rate and deliveries assisted by a skilled provider rose from 68% to 82% and 58% to 72% of all reported births, respectively. However, the rate of change for these outcomes was not substantially pronounced for the districts exposed to the RBF programme when compared with the enhanced financing study arm (C1), suggesting minimal impact of this programme for these outcomes over and above the changes seen in C1. For example, a substantial relative rise in the in-facility delivery rate by 12.8-percentage points in the RBF arm vis-à-vis the C2 (observational control) was lower than the relative gain in the enhanced financing arm (C1) which recorded a 17.5-percentage point relative gain vis-à-vis the C2 (observational control). The same basic pattern holds for a facility delivery assisted by a skilled provider – the RBF arm records a 10.1 percentage point gain in this measure while the C1 arm records a 14.2 percentage point gain. While the gains in C1 in these two measures are greater in magnitude than they are for the RBF arm, they are not significantly different from the RBF arm at standard levels.

Table 4: In-Facility delivery indicators

All births within two years of survey	Mean at baseline			Mean at endline			RBF vs. Control 1		RBF vs. Control 2		Control 1 vs. Control 2	
	RBF	Cont 1	Cont 2	RBF	Cont 1	Cont 2	Impact	p-value	Impact	p-value	Impact	p-value
Facility delivery	0.677	0.561	0.705	0.822	0.757	0.721	-0.049	0.308	0.128*	0.115	0.175*	0.038
Skilled provider and facility	0.583	0.489	0.620	0.719	0.668	0.656	-0.043	0.374	0.101	0.226	0.142*	0.144

Note: Sample size 4,488 births; * $p < 0.15$; linear probability model with difference-in-difference specification, including controls for district stratification. Errors are clustered at the PSU.

ANC Services

45. Table 5 provides summary indicators related to the receipt of antenatal care (ANC) and related services. Any ANC was already quite high at baseline in Zambia (96%-98%) and rises to near universal coverage by the endline study period (99% in all three areas). As a result, there are few observable gains in ANC coverage for RBF districts. There were no observable gains in the C1 study district for any of these measures. One key aspect of ANC coverage that the RBF pilot improved was the timing of the first ANC visit. At baseline, the mean timing of the first ANC visit occurred at the fourth month of pregnancy in all study areas. After two years of the RBF pilot, the timing for women in RBF areas improves to 3.8 months in RBF areas while increasing to around 4.2 months in both the C1 and C2 areas. This represents an improvement of almost two weeks in the timing of first ANC in the RBF arm as compared to the C1 and C2 arms.

Table 5: Antenatal care coverage

All pregnancies within two years	Mean at baseline			Mean at endline			RBF vs. Control 1		RBF vs. Control 2		Control 1 vs. Control 2	
	RBF	Cont 1	Cont 2	RBF	Cont 1	Cont 2	Impact	Exact	Impact	Exact	Impact	Exact
Any ANC	0.968	0.957	0.959	0.988	0.993	0.993	-0.015	0.266	-0.015	0.263	0.000	0.493
ANC: proper provider and facility	0.855	0.868	0.861	0.901	0.863	0.869	0.052	0.280	0.041	0.326	-0.014	0.442
4 or more ANC visits	0.701	0.662	0.624	0.702	0.666	0.656	-0.004	0.471	-0.034	0.293	-0.029	0.319
Number of ANC visits	4.204	4.015	3.980	4.106	3.991	4.067	-0.074	0.305	-0.187*	0.093	-0.112	0.218
Timing of first ANC visit	4.011	4.024	3.966	3.753	4.140	4.192	-0.372*	0.096	-0.476*	0.043	-0.108	0.358

Note: Sample size 4,543 pregnancies; * $p < 0.15$; For binary measures, the linear probability model is used, including controls for district stratification and errors are clustered at the PSU. For number of ANCs and no. months pregnant at first ANC, standard OLS regression with controls for district stratification. Errors are clustered at the PSU.

Postnatal Care (PNC)

46. There were broad-based relative gains in PNC coverage, as summarized in Table 6, most likely driven to a large degree by the gains in the in-facility delivery rate. However, the rate of change (for any PNC) was more rapid in C1 vs C2 (a 13 percentage point increase) as compared to RBF vs C2. While absolute gains in PNC are precisely estimated for C1 vs C2 this relative difference in effectiveness between RBF vs C1, and RBF vs C2 is not precisely estimated.

Table 6: Postnatal care coverage

All births within two years of survey	Mean at baseline			Mean at endline			RBF vs. Control 1		RBF vs. Control 2		Control 1 vs. Control 2	
	RBF	Cont 1	Cont 2	RBF	Cont 1	Cont 2	Impact	Exact	Impact	Exact	Impact	Exact
Any PNC	0.692	0.557	0.758	0.836	0.753	0.819	-0.051	0.340	0.082	0.249	0.132*	0.129
PNC: proper provider and facility	0.608	0.476	0.652	0.770	0.651	0.712	-0.010	0.473	0.103	0.238	0.112	0.219

Note: Sample size 4,488 births; * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$; linear probability model with difference-in-difference specification, including controls for district stratification. Errors are clustered at the PSU.

Contraceptive Use

47. With regard to increased use of modern contraceptive methods among women of reproductive age, Table 7 explores the programme effectiveness among the more restricted sample of women who reside in household that have reported a birth in the previous 2 years before survey (i.e. those women targeted for interview in the household surveys). The results show that the use of modern or any contraceptive method remained relatively unchanged at endline as compared to the baseline across the 3 study arms, but these results were statistically insignificant. However, there were more family planning (FP) outreach services being conducted in the RBF districts vis-à-vis the C2 districts but even greater in C1 districts when compared to both RBF and C2 districts. Note that this population is a non-random subset of all women on reproductive age and thus may not represent population changes in family planning usage as a result of RBF or C1 activities.

Table 7: Family planning indicators

Women 15-49 in recent birth hhs	Mean at baseline			Mean at endline			RBF vs. Control 1		RBF vs. Control 2		Control 1 vs. Control 2	
	RBF	Cont 1	Cont 2	RBF	Cont 1	Cont 2	Impact	Exact	Impact	Exact	Impact	Exact
Any Contraception	0.568	0.560	0.612	0.601	0.619	0.688	-0.024	0.380	-0.039	0.307	-0.017	0.416
Modern Contraception (of FP users)	0.921	0.851	0.925	0.910	0.886	0.908	-0.045	0.204	0.002	0.486	0.046	0.195
Any FP outreach	0.772	0.697	0.847	0.881	0.884	0.871	-0.078*	0.127	0.083*	0.110	0.159*	0.005

Note: Sample size 5,032 women aged 15-49 in households with a recent birth; * p<0.15; Linear probability model with difference-in-difference specification, including controls for district stratification. Errors are clustered at the PSU.

Immunization

48. Table 8 shows immunization coverage by different vaccines and timing of the vaccination across the 3 study arms. The results show that full immunization coverage across all the 3 study areas declined, but to a lesser degree in RBF districts suggesting that RBF may have been somewhat protective. However, the relative protective effects are not precisely estimated. For some measures of immunization coverage – BCG and DPT vaccines – RBF districts performed significantly better than the C2 districts. On the other hand, C1 performed better than C2 districts on any vaccination coverage.

Table 8: Immunization Coverage for children aged 12-23 months

Children between 12 and 23 months	Mean at baseline			Mean at endline			RBF vs. Control 1		RBF vs. Control 2		Control 1 vs. Control 2	
	RBF	Cont 1	Cont 2	RBF	Cont 1	Cont 2	Impact	Exact	Impact	Exact	Impact	Exact
Fully vaccinated	0.395	0.291	0.322	0.383	0.180	0.216	0.116	0.271	0.052	0.391	-0.046	0.405
Any vaccinations	0.944	0.883	0.922	0.955	0.962	0.921	-0.066	0.171	0.015	0.418	0.081*	0.115
BCG inject ever received	0.956	0.978	0.976	1.000	0.995	0.957	0.031	0.232	0.07*	0.027	0.028	0.252
Polio ever received	0.956	0.978	0.964	0.966	0.985	0.957	0.010	0.423	0.027	0.273	0.014	0.379
Polio received right after birth	0.288	0.291	0.325	0.392	0.368	0.330	0.003	0.484	0.067	0.317	0.062	0.331
DPT ever received	0.971	0.952	0.958	0.986	0.975	0.919	-0.010	0.418	0.061*	0.082	0.056*	0.104
DPT freq given	2.669	2.446	2.559	2.752	2.452	2.462	0.098	0.354	0.168	0.259	0.083	0.374
Measles ever received	0.928	0.898	0.900	0.914	0.920	0.921	-0.053	0.231	-0.033	0.320	0.013	0.430

Note: Sample size 768 children between 12 and 23 months old; * p<0.15; Linear probability model with difference-in-difference specification, including controls for district stratification. Errors are clustered at the PSU.

Out-patient utilization

49. One incentivized RBF service was general curative care, which was given the lowest unit-price. Table 9 explores health seeking patterns in the population conditional on report of illness in the four weeks before survey. While health seeking behaviour is relatively high both at baseline and endline, there is no apparent differential trend in general health-seeking behaviour across study arms. There is some indication that for under-5s there is a relative change towards Community

Health Workers (CHWs) in both RBF and C1 districts, although this effect is largely driven by a decline of CHW health-seeking in C2 areas.

Table 9: Health seeking behaviour for general illness, separately for under-5s and over-5s

	<i>Mean at baseline</i>			<i>Mean at endline</i>			<i>RBF vs. Control 1</i>		<i>RBF vs. Control 2</i>		<i>Control 1 vs. Control 2</i>	
	RBF	Cont 1	Cont 2	RBF	Cont 1	Cont 2	Impact	Exact	Impact	Exact	Impact	Exact
All children under 5												
Any sickness in previous 4 weeks	0.185	0.123	0.148	0.172	0.099	0.142	0.013	0.425	-0.007	0.461	-0.023	0.372
Sought advice for illness	0.735	0.751	0.849	0.768	0.785	0.902	0.052	0.188	-0.018	0.383	-0.065	0.134
Visited gov facility for illness	0.928	0.930	0.854	0.929	0.926	0.932	-0.029	0.286	-0.095**	0.027	-0.035	0.252
Visited any facility for illness	0.977	0.962	0.891	0.945	0.941	0.961	-0.030	0.238	-0.098**	0.002	-0.048	0.126
Visited CHW for illness	0.009	0.019	0.063	0.004	0.059	0.005	-0.009	0.426	0.054**	0.022	0.057	0.013
HH members over 5												
Any sickness in previous 4 weeks	0.072	0.058	0.063	0.050	0.046	0.063	-0.010	0.340	-0.022	0.175	-0.012	0.307
Sought advice for illness	0.668	0.641	0.749	0.649	0.717	0.761	-0.038	0.346	-0.046	0.315	-0.009	0.464
Visited gov facility for illness	0.926	0.959	0.869	0.940	0.878	0.824	0.09*	0.084	0.037	0.303	-0.025	0.366
Visited any facility for illness	0.974	0.979	0.915	0.953	0.885	0.857	0.082*	0.095	0.039	0.281	-0.021	0.379
Visited CHW for illness	0.011	0.000	0.045	0.000	0.050	0.019	-0.054*	0.001	0.006	0.463	0.043	0.047

Note: Sample size 6,981 U5s and 17,059 over 5s. * p<0.15; Linear probability model with difference-in-difference specification, including controls for district stratification. Errors are clustered at the PSU.

5.2 Quality of services

5.2.1 Structural Quality

50. This section evaluates the status of infrastructure, and availability of essential drugs and equipment at health facilities. The data was derived from the two rounds of the facility survey conducted in RBF and control facilities at baseline and endline. The impact estimate given is the relative change in indicators for the RBF facilities compared with the C1 and C2 facilities between baseline and endline surveys, controlling for district level stratification variables.

a) Status of Infrastructure

The status of infrastructure at the facilities was assessed through direct observation. Relevant dimensions of infrastructure included availability of power, water, tele-communication systems, disinfectants, an outpatient consultation room, availability of key elements in the outpatient room for optimal service delivery, and provision of biomedical waste disposal. An infrastructure index was constructed, which included the following equally-weighted items: continuous availability of power, water, communication and disinfectants, provision of sharp disposal, and basin with soap and water in outpatient room. As shown in Table 10, within the individual infrastructure measures (variables), the results were statistically insignificant but the infrastructure index showed higher gains in the RBF facilities as compared to C2.

Table 10: Status of infrastructure

	<i>Mean at baseline</i>			<i>Mean at endline</i>			RBF vs. Control 1		RBF vs. Control 2	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
Facility has electric power	0.82	0.72	0.78	0.97	0.91	0.88	-0.024	0.788	0.055	0.618
Facility experiences no power outage	0.45	0.39	0.50	0.79	0.77	0.67	-0.019	0.881	0.194	0.159
Facility experiences no water outage	0.79	0.79	0.88	0.88	0.84	0.91	0.041	0.688	0.051	0.476
Facility has functioning two-way radio	0.35	0.28	0.41	0.42	0.42	0.35	-0.026	0.836	0.141	0.203
Facility has phone line	0.34	0.26	0.22	0.47	0.40	0.42	-0.012	0.911	-0.077	0.591
Facility has patient transportation means	0.58	0.53	0.70	0.69	0.69	0.67	-0.022	0.875	0.134	0.291
Facility has functioning computer	0.22	0.32	0.32	0.34	0.29	0.29	0.126	0.390	0.144	0.372
Facility has general outpatient consultation room	0.86	0.87	0.85	0.95	0.96	0.90	0.004	0.948	0.050	0.623
Facility experiences no stock-out of disinfectant	0.79	0.83	0.80	0.92	0.92	0.85	0.052	0.542	0.098	0.238
Facility has functioning incinerator for medical waste	0.31	0.48	0.43	0.45	0.55	0.48	0.107	0.282	0.092	0.193
Infrastructure index	-0.09	0.01	0.14	0.07	0.02	-0.15	0.195	0.470	0.483*	0.099

Note: Results from 348 facilities; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

b) Availability of Drugs

Table 11 shows availability of tracer drugs at health facilities 30 days prior to the day of the survey. Tracer drugs included general antibiotics, analgesics, family planning, anti-malarials, anti-tuberculosis, antiretroviral, emergency obstetric care (EMOC), vaccines, diagnostic kits, fluids, and electrolytes. A drug availability index was constructed assigning equal weights to the individual items and was further standardized. The items were – Tetracycline eye ointment, Amoxicillin, Paracetamol, Cotrimoxazole, Iron and Folic acid, Vitamin A, and ORS. The results were statistically insignificant except for ACT and Amocixillin tablets. Availability of ACT drugs increased by 27 percentage points in the RBF facilities as compared to the C2 facilities but availability of Amocixillin tablets decreased by 21 percentage points in the RBF facilities as compared to the C1 facilities.

Table 11: Availability of drugs

	<i>Mean at baseline</i>			<i>Mean at endline</i>			RBF vs. Control 1		RBF vs. Control 2	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
Paracetamol tabs	0.86	0.86	0.86	0.82	0.91	0.87	-0.08	0.289	-0.04	0.699
Amoxicillin tabs	0.89	0.83	0.91	0.77	0.92	0.73	-0.21*	0.075	0.05	0.607
Iron tabs	0.85	0.81	0.74	0.92	0.92	0.87	-0.03	0.722	-0.03	0.824
Folic acid tabs	0.82	0.77	0.88	0.82	0.87	0.77	-0.09	0.455	0.13	0.259
Cotrimoxazole	0.84	0.79	0.91	0.79	0.79	0.74	-0.07	0.486	0.11	0.337
Vitamin A	0.77	0.78	0.77	0.64	0.70	0.67	-0.04	0.774	-0.01	0.957
Oral contraceptive pills	0.93	0.83	0.81	0.83	0.63	0.63	0.07	0.417	0.07	0.634
Implant	0.72	0.74	0.78	0.72	0.71	0.77	0.02	0.916	0.02	0.919
Artemisinin-Based Combination Therapy (ACT)	0.79	0.75	0.93	0.96	0.88	0.82	0.04	0.693	0.27***	0.008
Rifampicin	0.70	0.70	0.59	0.73	0.66	0.64	0.04	0.828	-0.03	0.894
Magnesium sulfate	0.52	0.58	0.57	0.53	0.67	0.67	-0.08	0.635	-0.06	0.768
Misoprostol	0.65	0.40	0.50	0.32	0.36	0.40	-0.30	0.278	-0.20	0.365
Oxytocin	0.75	0.83	0.85	0.83	0.98	0.90	-0.07	0.525	0.05	0.656
Pentavalent vaccines	0.79	0.86	0.82	0.81	0.70	0.67	0.17	0.185	0.18	0.155
Malaria rapid diagnostic kits	0.83	0.84	0.83	0.82	0.86	0.77	-0.04	0.742	0.05	0.766
HIV testing kit	0.84	0.87	0.88	0.81	0.73	0.81	0.10	0.448	0.04	0.734
Pregnancy testing kit	0.54	0.54	0.61	0.36	0.51	0.51	-0.16	0.501	-0.07	0.710
Urine Dipstick	0.65	0.52	0.68	0.61	0.70	0.67	-0.22	0.336	-0.01	0.971
Drug availability index	0.01	-0.05	0.03	-0.01	0.04	-0.01	-0.08	0.844	0.06	0.893

Note: Results from 348 facilities; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

c) Availability of medical equipment

Table 12 below shows an increase in the availability of medical equipment in RBF health facilities as compared to C2 health facilities as observed through an increase in the overall equipment index. Specific gains in the RBF facilities as compared to the C2 facilities were observed in the availability of infant weighing scales (22 percentage points), forceps (16 percentage points), and needle holders (25 percentage points). There was also a higher availability of tape measures in the RBF facilities vs. C1 facilities by 15 percentage points but no overall gain was observed in the medical equipment index between the RBF and C1 health facilities.¹⁰ The medical equipment index did show a relative increase in RBF areas vis-à-vis C2 facilities.

¹⁰ All aggregated indices in this report, unless noted otherwise, are normalized with mean zero and standard deviation of 1, with aggregation weights determined by principal components analysis.

Table 12: Availability of medical equipment

	<i>Mean at baseline</i>			<i>Mean at endline</i>			RBF vs. Control 1		RBF vs. Control 2	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
Children's weighing scale	0.94	0.98	0.89	0.03	0.06	0.04	0.01	0.807	-0.05	0.433
Height measure	0.47	0.51	0.59	0.05	0.04	0.02	0.07	0.496	0.15	0.211
Tape measure	0.52	0.68	0.61	0.04	0.04	0.02	0.15*	0.097	0.11	0.399
Adult weighing scale	0.92	0.89	0.91	0.07	0.08	0.10	-0.02	0.759	-0.02	0.790
Blood pressure instrument	0.81	0.74	0.83	0.03	0.04	0.04	-0.07	0.447	0.02	0.822
Thermometer	0.90	0.89	0.85	0.04	0.09	0.00	-0.06	0.442	-0.01	0.849
Stethoscope	0.83	0.89	0.85	0.05	0.02	0.08	0.09	0.280	-0.01	0.911
Fetoscope	0.81	0.81	0.83	0.05	0.04	0.04	0.00	0.967	0.03	0.689
Otoscope	0.20	0.19	0.24	0.00	0.00	0.00	-0.01	0.878	0.04	0.680
Electric autoclave (pressure and wet heat)	0.07	0.13	0.13	0.03	0.04	0.04	0.04	0.506	0.05	0.313
Refrigerator	0.77	0.72	0.70	0.18	0.23	0.27	-0.08	0.481	-0.15	0.232
Delivery table/bed	0.67	0.66	0.76	0.00	0.00	0.00	-0.01	0.890	0.09	0.396
Partograph	0.48	0.53	0.57	0.01	0.04	0.00	0.02	0.875	0.10	0.461
Baby scale (infant weighing scale)	0.69	0.72	0.89	0.01	0.00	0.00	0.05	0.643	0.22***	0.007
Forceps, artery	0.67	0.74	0.83	0.01	0.00	0.00	0.08	0.406	0.16**	0.011
Needle holder	0.60	0.55	0.85	0.00	0.04	0.00	-0.09	0.389	0.25***	0.001
Bag Valve Mask (Ambu bag)	0.20	0.21	0.26	0.03	0.02	0.04	0.02	0.847	0.05	0.694
Guedel airways-neonatal, child, and adult	0.12	0.17	0.09	0.00	0.00	0.00	0.05	0.426	-0.03	0.564
Equipment availability index	-0.05	-0.04	0.12	0.04	0.03	-0.11	0.03	0.917	0.37*	0.088

Note: Results from 348 facilities; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

d) Mapping the quality checklist at health facilities

As described above, during the RBF implementation period, quality assessments were being conducted by district hospitals at all health facilities on a quarterly basis using a standard quality checklist which had 10 dimensions on quality improvement. The quality checklist which was being used by the district hospitals and the health facility survey instrument for this IE overlap on several items. These common items were extracted from the health facility survey instrument and the weights from the quality checklist were used to construct a quality index. Standardized indices were constructed for each quality dimension. Table A3 in Appendix 3 maps out the common variables in both instruments. Table 13 reveals that the quality of delivery rooms in RBF facilities was better than the delivery rooms in C1 and C2 districts. In addition, the quality of curative care in RBF facilities was better than in the C2 facilities.

Table 13: Mapping the quality checklist at health facilities

	<i>Mean at baseline</i>			<i>Mean at endline</i>			<i>RBF vs. Control 1</i>		<i>RBF vs. Control 2</i>	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
Curative Care	-0.03	-0.01	0.06	0.22	-0.16	-0.24	0.39	0.204	0.28**	0.042
Family planning	0.03	-0.03	-0.02	0.10	-0.10	-0.10	0.15	0.578	0.08	0.546
Delivery Room	-0.11	-0.04	0.23	0.36	-0.19	-0.45	0.61**	0.010	0.57***	0.000
Supply management	-0.02	-0.10	0.13	0.05	-0.25	0.15	0.22	0.496	0.02	0.847

Note: Results from 348 facilities; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

5.2.2 Process Quality

51. This section summarizes the relative impacts of RBF on the quality of care provided for antenatal, postnatal and child health care services. There are a number of sources for this data: direct clinical observations of consultations, exit interviews administered to patients as they are leaving their consultations and recall data of procedures received from the household survey. Based on these interviews, client satisfaction was also estimated.

a) Knowledge of maternal health services

The household survey also measured mother's knowledge of danger signs during pregnancy. The knowledge of individual danger signs, conveyed in Table 14, can be affected through contact with the health system either at the communal or facility level, including outreach programs. In terms of effectiveness of such programs and contact, at least with regard to maternal knowledge, RBF districts appear to be somewhat effective. Women residing in RBF districts are significantly more likely to list three out of the 12 danger signs such as severe pain or foetal stillness when compared with C1 districts. Results when compared with C2 are similar.

Table 14: Knowledge of maternal health danger signs: Results from the household survey

Women 15-49 in recent birth hhs	<i>Mean at baseline</i>			<i>Mean at endline</i>			<i>RBF vs. Control 1</i>		<i>RBF vs. Control 2</i>		<i>Control 1 vs. Control 2</i>	
	RBF	Cont 1	Cont 2	RBF	Cont 1	Cont 2	Impact	Exact	Impact	Exact	Impact	Exact
Excessive Vaginal Bleeding	0.458	0.532	0.471	0.639	0.630	0.617	0.080	0.237	0.045	0.343	-0.037	0.373
Fever	0.232	0.305	0.276	0.318	0.283	0.372	0.108	0.186	-0.015	0.450	-0.120	0.159
Swollen face, hands or legs	0.224	0.290	0.209	0.249	0.291	0.284	0.023	0.372	-0.055	0.218	-0.082*	0.121
Difficulty breathing	0.137	0.144	0.126	0.086	0.062	0.069	0.030	0.270	0.005	0.456	-0.022	0.328
Sever headache	0.131	0.174	0.158	0.138	0.166	0.201	0.018	0.383	-0.041	0.245	-0.058	0.169
Convulsions / Fits	0.057	0.062	0.040	0.022	0.030	0.040	-0.002	0.478	-0.035	0.200	-0.033	0.209
Lightheadedness / Dizziness	0.149	0.144	0.155	0.164	0.104	0.157	0.056	0.158	0.014	0.412	-0.049	0.194
Blurred vision	0.050	0.056	0.050	0.059	0.045	0.043	0.019	0.241	0.013	0.313	-0.005	0.427
High blood pressure	0.171	0.188	0.144	0.232	0.177	0.138	0.071	0.165	0.070	0.166	0.004	0.482
Severe pain in lower belly	0.213	0.229	0.321	0.283	0.137	0.230	0.164*	0.039	0.162*	0.041	-0.001	0.498
Water leaks or breaks	0.108	0.141	0.125	0.202	0.100	0.101	0.134*	0.043	0.12*	0.065	-0.018	0.415
Baby stops or reduces moving	0.140	0.210	0.134	0.235	0.224	0.157	0.078*	0.129	0.071	0.154	-0.009	0.453

Note: 5,241 women of reproductive age. * p<0.15; Linear probability model with difference-in-difference specification, including controls for district stratification. Errors are clustered at the PSU.

b) Maternal care

Table 15 provides recall information from the household survey on ANC processes received while attending ANC services. The results show some gains in process quality of ANC in the RBF districts in

some of the indicators assessed. This is for provision of a tetanus injection (vs C1); and any iron and malaria drugs taken which exhibited a higher relative gain in RBF communities than C2 communities. In most of the other indicators, there was a relative decline in RBF communities as compared to C1 and C2 communities which includes: receiving a urine test during pregnancy, number of days of iron supplementation by pregnant women, and testing of blood (vs C1). On the other hand, blood testing and any iron taken shows relatively higher gains in C1 communities when compared with C2 communities (although any tetanus is lower).

Table 15: Process quality of antenatal care provided: Results from the household survey

All pregnancies within 2 years of int.	Mean at baseline			Mean at endline			RBF vs. Control 1		RBF vs. Control 2		Control 1 vs. Control 2	
	RBF	Cont 1	Cont 2	RBF	Cont 1	Cont 2	Impact	Exact	Impact	Exact	Impact	Exact
Blood pressure measured	0.863	0.786	0.833	0.937	0.889	0.933	-0.029	0.284	-0.028	0.294	0.000	0.498
Urine sample taken	0.197	0.179	0.234	0.201	0.312	0.334	-0.129*	0.064	-0.095*	0.138	0.031	0.363
Blood sample taken	0.801	0.762	0.889	0.870	0.930	0.901	-0.099	0.161	0.057	0.325	0.153*	0.031
Any tetanus injection	0.803	0.823	0.807	0.858	0.819	0.900	0.06*	0.123	-0.039	0.233	-0.100*	0.018
Number of tetanus injections	1.567	1.481	1.525	1.405	1.350	1.299	-0.041	0.384	0.067	0.307	0.109	0.201
Any iron taken	0.872	0.883	0.930	0.960	0.973	0.974	-0.001	0.494	0.044*	0.123	0.045*	0.119
Number of days iron taken	44.509	35.685	39.759	63.836	71.004	74.649	-16.56*	0.144	-15.300	0.165	0.453	0.488
Malaria drugs taken	0.909	0.916	0.943	0.981	0.963	0.979	0.026	0.234	0.036*	0.149	0.010	0.392

Note: Sample size 4,256 births; * p<0.15; Linear probability model with difference-in-difference specification, impact estimates adjusted for district pair matching with standard errors clustered at district level. For number of tetanus injections and number of days iron taken, OLS with same specification.

Table 16 presents similar quality measures for ANC, but measured using patient exit interviews conducted as patients left health facilities during the health facility survey. Participants were asked a series of questions as to whether certain ANC services were performed during their visit. The items were summed to create a composite score for the number of ANC services that were performed and then converted to z-scores. Finally, a dichotomous variable was created to measure if all of the ANC services were completed for each participant. This variable was standardized before use in the final analyses. The results show that more women reported to have received advice on diet (14 percentage points) in RBF as compared to C1 facilities. In addition, compared to C2 facilities, more women in RBF facilities reported to have had their abdomen measured (9 percentage points) and palpated (12 percentage points). On the other hand, women who attended C1 facilities received explanations on the side effects of iron folic acid tablets (15 percentage points) as compared to those who went to RBF facilities.

Table 16: Process quality of antenatal care: Results from patient exit interviews

	<i>Mean at baseline</i>			<i>Mean at endline</i>			<i>RBF vs. Control 1</i>		<i>RBF vs. Control 2</i>	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
Weighed	0.85	0.82	0.92	0.9	0.89	0.91	-0.02	0.632	0.06	0.251
Blood pressure measured	0.61	0.58	0.72	0.84	0.84	0.87	-0.03	0.809	0.08	0.452
Urine sample collected	0.15	0.16	0.18	0.19	0.2	0.32	0.01	0.928	-0.09	0.298
Blood sample collected	0.56	0.53	0.66	0.54	0.5	0.66	0	0.987	-0.02	0.835
Abdomen measured	0.84	0.92	0.94	0.95	0.95	0.95	0.07	0.152	0.09*	0.063
Abdomen palpated	0.39	0.41	0.48	0.49	0.49	0.45	0	0.987	0.12*	0.083
Advice on diet	0.51	0.66	0.58	0.72	0.72	0.77	0.14***	0.009	0.02	0.85
Given/prescribed iron folic acid	0.84	0.9	0.83	0.95	0.96	0.92	0.05	0.149	0.02	0.835
Side effects of iron folic acid explained	0.24	0.12	0.18	0.14	0.18	0.27	-0.15**	0.042	-0.18	0.133
Given/prescribed antimalarials	0.71	0.62	0.74	0.81	0.82	0.84	-0.09	0.421	0	0.962
Explained danger signs of pregnancy	0.55	0.57	0.62	0.78	0.83	0.84	-0.04	0.591	0	0.99
Quality of ANC index	-0.08	-0.05	0.27	0	-0.01	0.01	0.02	0.921	0.33	0.165

Sample size 1,954 clients; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

Table 17 presents results on the process quality of postnatal coverage. The RBF appears to have had no positive effect on the process quality of postnatal care, except that rates of immediate breastfeeding are significantly higher when compared with C2. The results show a relatively larger gain in C1 districts both in terms of Vitamin A given soon after delivery and in the immediate initiation of breastfeeding when compared with RBF and, especially, C2 districts.

Table 17: Process quality of postnatal care provided: Results from the household survey

	<i>Mean at baseline</i>			<i>Mean at endline</i>			<i>RBF vs. Control 1</i>		<i>RBF vs. Control 2</i>		<i>Control 1 vs. Control 2</i>	
	RBF	Cont 1	Cont 2	RBF	Cont 1	Cont 2	Impact	Exact	Impact	Exact	Impact	Exact
All births within 2 years of survey	0.412	0.348	0.508	0.476	0.506	0.549	-0.091	0.177	0.025	0.404	0.115*	0.113
Vit A given after delivery	0.979	0.985	0.985	0.976	0.982	0.976	0.000	0.499	0.006	0.364	0.006	0.357
Ever breastfed infant	0.655	0.570	0.736	0.804	0.787	0.749	-0.070	0.288	0.136*	0.136	0.206*	0.041

Note: Sample size 4,252 births; * p<0.15; Linear probability model with difference-in-difference specification, including province-level controls. Errors are clustered at the district level

c) Child health care

Quality of child care was measured through patient exit interviews (Table 18). Participants were asked a series of questions as to whether certain child care services were conducted during the visit. The six items were summed to create a composite score for the number of child care services that were performed. Finally, a dichotomous variable was created to measure if all five of the child care services were completed. This variable was standardized before use in final analyses. The data for all the variables (except for plotted on a growth chart – RBF vs C1) are statistically insignificant suggesting that there are no differences in the quality of child health care as determined through exit interviews.

Table 18: Process quality of child health care: Results from exit interviews

	<i>Mean at baseline</i>			<i>Mean at endline</i>			<i>RBF vs. Control 1</i>		<i>RBF vs. Control 2</i>	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
Asked age	0.88	0.91	0.88	0.91	0.95	0.90	-0.01	0.880	0.02	0.776
Weighed child	0.62	0.63	0.61	0.43	0.51	0.37	-0.07	0.378	0.06	0.498
Measured height	0.08	0.06	0.07	0.05	0.13	0.07	-0.10	0.104	-0.02	0.577
Plotted a growth chart	0.37	0.35	0.41	0.12	0.23	0.17	-0.13*	0.080	-0.02	0.800
Physically examined	0.62	0.60	0.64	0.60	0.67	0.69	-0.09	0.327	-0.08	0.350
Quality of care index	-0.02	-0.03	0.07	-0.01	0.08	-0.06	-0.09	0.669	0.14	0.565

Sample size 2,197 clients; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

d) Client satisfaction

This section summarizes the effect of the intervention on client satisfaction in antenatal care and child health care as measured through patient exit interviews. Participants were asked to rate whether the overall quality of the services was satisfactory on a Likert scale (1=strongly disagree to 5=strongly agree). Table 19 shows that most of the results were statistically insignificant but RBF health facilities had relatively more ANC clients who reported that health workers spent sufficiently more time during consultations with patients as compared to both C1 and C2 health facilities. There was also more trust in health workers operating in RBF facilities as compared with C1 for both ANC and child care services (Tables 19 and 20).

Table 19: Client satisfaction in antenatal care: Results from exit interviews

	<i>Mean at baseline</i>			<i>Mean at endline</i>			<i>RBF vs. Control 1</i>		<i>RBF vs. Control 2</i>	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
The health facility is clean	0.77	0.81	0.72	0.90	0.93	0.90	0.01	0.843	-0.04	0.448
The health staff are courteous and respectful	0.82	0.90	0.87	0.88	0.94	0.95	0.02	0.542	-0.02	0.731
The amount of time you spent waiting to be seen by a health provider was reasonable	0.70	0.73	0.74	0.79	0.78	0.73	0.02	0.747	0.09	0.198
The health worker spent a sufficient amount of time with the patient	0.82	0.92	0.90	0.94	0.96	0.95	0.08*	0.067	0.08*	0.081
The hours the facility is open is adequate to meet the needs of the community	0.76	0.83	0.82	0.92	0.94	0.93	0.05	0.327	0.06	0.340
All in all, you trust the health worker completely in this health facility	0.78	0.89	0.84	0.94	0.97	0.96	0.07*	0.066	0.03	0.569
Satisfaction index	-0.11	0.13	0.09	-0.07	0.12	0.02	0.04	0.826	0.12	0.574

Sample size 1,954 clients; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

Table 20: Client satisfaction on child health care: Results from exit interviews

	<i>Mean at baseline</i>			<i>Mean at endline</i>			RBF vs. Control 1		RBF vs. Control 2	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
The health facility is clean	0.77	0.81	0.79	0.86	0.87	0.90	0.03	0.657	-0.03	0.732
The health staff are courteous and respectful	0.77	0.87	0.84	0.86	0.90	0.85	0.04	0.445	0.06	0.305
The amount of time you spent waiting to be seen by a health provider was reasonable	0.71	0.75	0.72	0.71	0.77	0.77	-0.02	0.823	-0.06	0.477
The health worker spent a sufficient amount of time with the patient	0.80	0.90	0.87	0.89	0.93	0.89	0.07	0.128	0.07	0.234
The hours the facility is open is adequate to meet the needs of the community	0.75	0.85	0.83	0.89	0.91	0.92	0.08	0.157	0.04	0.502
All in all, you trust the health worker completely in this health facility	0.77	0.88	0.85	0.91	0.91	0.94	0.11*	0.057	0.04	0.504
Satisfaction index	-0.12	0.15	0.09	-0.09	0.09	0.07	0.09	0.617	0.04	0.858

Sample size 2,197 clients; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

5.3 Effect of RBF and enhanced financing on health system performance measures

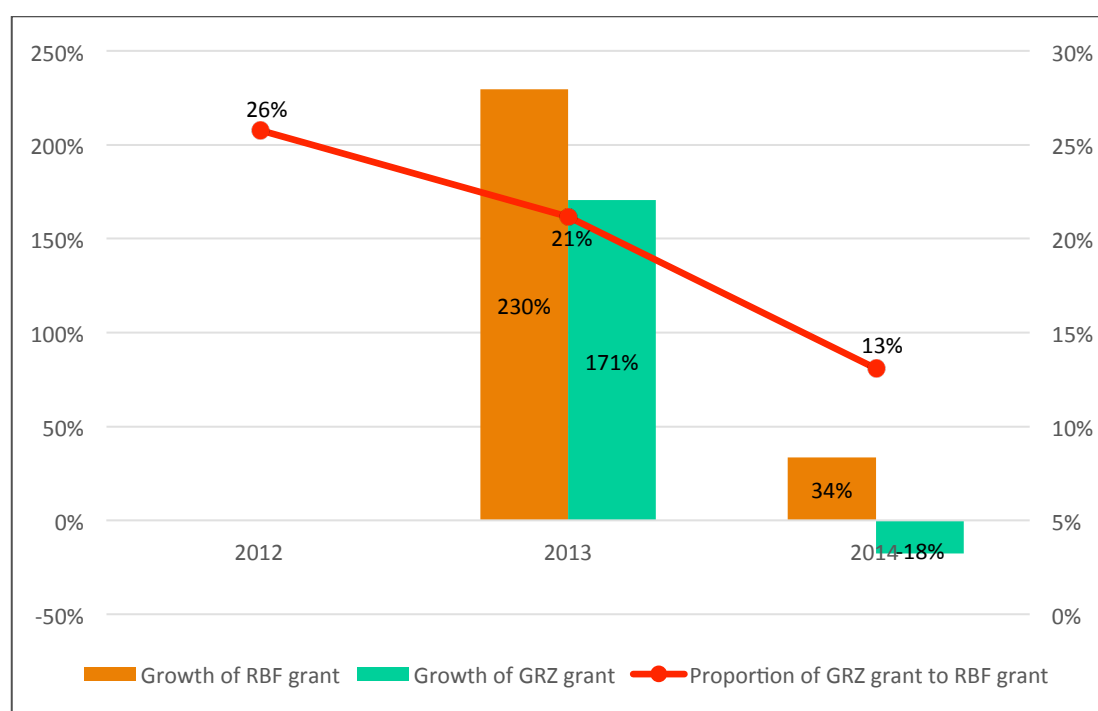
5.3.1 Level of Revenue in RBF Health Facilities, GRZ vs RBF Grants

52. Figure 5 compares the proportion of the GRZ grant¹¹ to the RBF grant in the RBF health facilities during the RBF implementation period. The idea was to track the growth of the RBF grant against the GRZ grant and to establish whether there was any aid fungibility¹² or substitution of financing. The data shows that in proportion to the RBF grant, the GRZ grant declined by half from 26% in 2012 to 13% in 2014. The data further shows that the RBF grant was growing faster than the GRZ grant. The RBF grant grew by 230% between 2012 and 2013, and by 34% between 2013 and 2014. Meanwhile, the GRZ grant grew by 171% between 2012 and 2013, and declined by 18% between 2013 and 2014.

¹¹ The Zambian government provides operational grants to all health facilities on a monthly basis. These grants are allocated on the basis of the health facility's catchment population and can be used for recurrent operational activities, outreach, and purchase of goods and services including meal allowances,

¹² Aid fungibility is when donor funding for health substitutes for—rather than complements—health financing by recipient governments.

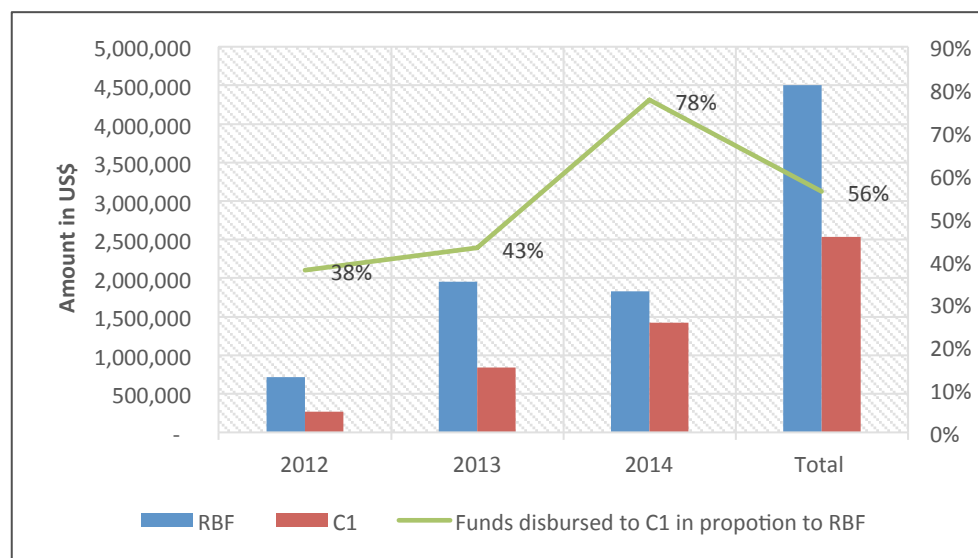
Figure 5: Proportion of GRZ grant to RBF grant



5.3.2 Level of RBF revenue, RBF vs C1

53. The study was designed to equalize total RBF financing between health facilities in RBF districts and C1 districts to rule out the “money” effect. Actual RBF financing was tracked by the study and found divergences between intended and actual financing received in the C1 arm. These divergences were largely due to the available disbursement mechanisms used to reach C1 facilities. C1 districts were using a financing system where the RBF matching grant was being channelled through the District Medical Office (DMO) for further disbursement to health facilities under each district, unlike in the RBF districts which did not use the DMO as a channel. In addition, money disbursed to the health facilities in C1 districts was required to be retired before replenishment. This was contrary to health facilities in the RBF arm where RBF payments were disbursed directly in the health centre bank accounts and no retirements were required.
54. As a consequence, health facilities in the C1 districts did not receive the same amounts as the RBF districts due to delayed retirement and low absorptive capacity. As such, the input financing (C1) arm only received 38% in 2012, 43% in 2013, and 78% in 2014 as compared to the RBF arm. The overall disbursement for all C1 districts was 56% (from April 2012 to October 2014) and thus not at par with the RBF arm as expected. This is shown on Figure 6 below. This divergence from the intended design is important to note as results on the relative effectiveness and cost-effectiveness between the study arms are discussed.

Figure 6: Funds disbursed to C1 districts as compared to RBF districts



Source: Authors' construction from RBF operational data

5.3.3 Allocation and use of RBF funds, RBF vs C1

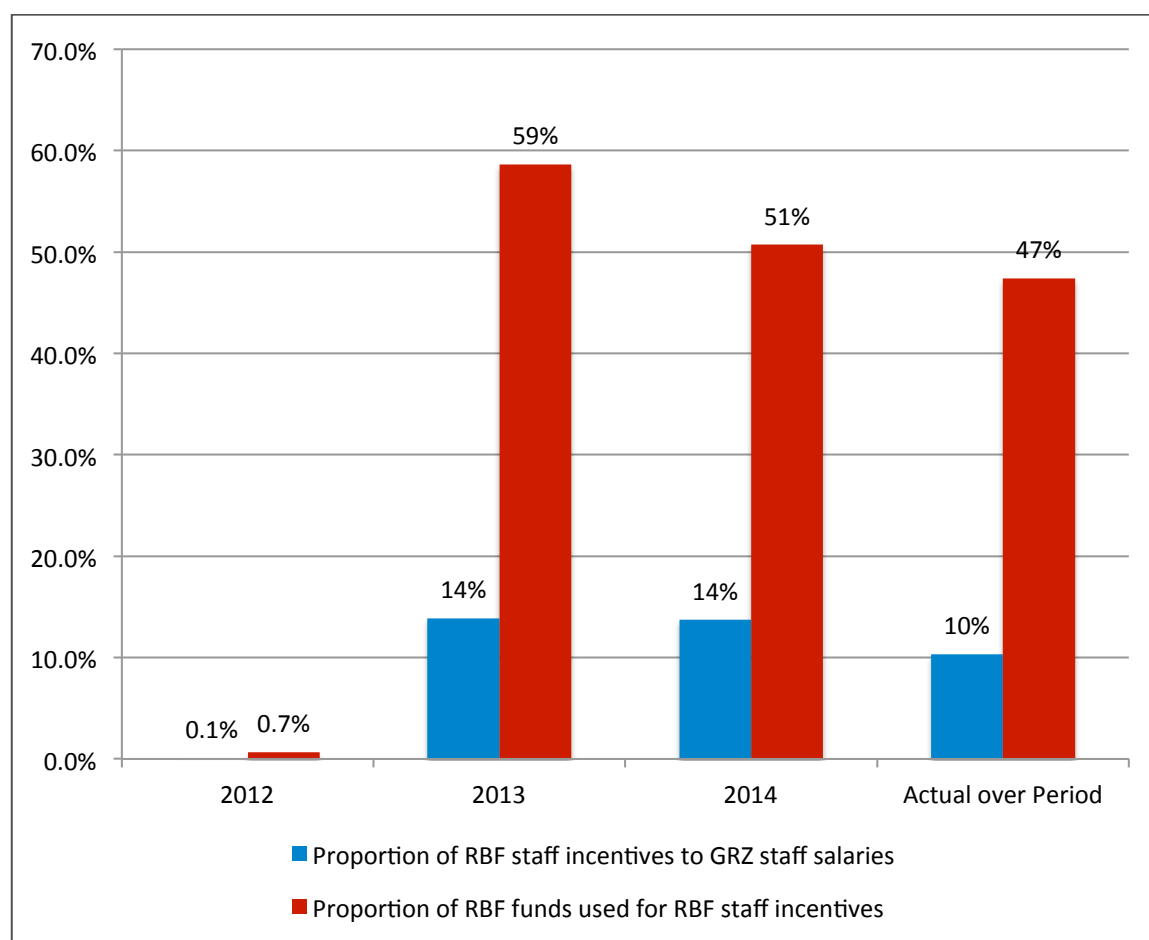
a) Allocation and use of the RBF performance grants in RBF Health Facilities

In order to help interpret the findings, it is useful to review study information on the use of funds in the RBF and C1 districts at health facility level. Health facilities in the RBF districts were allowed to use a maximum of 60% of their RBF funds for staff incentives, and a minimum of 40% for investments and other recurrent/operational costs at the health facilities and communities. In reality, the percentages allocated varied by health facility, both across districts and over time. For example, in 2012, almost all the health facilities didn't spend on staff incentives which explains why the proportion of RBF funds used for RBF staff incentives was only 0.7%, and the proportion of RBF staff incentives to GRZ staff salaries 0.1% (Figure 7). Overall, health facilities in the RBF intervention group allocated 47% of the total RBF funds for staff incentives, and 53% for investments between April 2012 and October 2014, the full Zambia HRBF implementation period (Figure 7).

The 53% which was spent on investment in the RBF districts was generally used for the purchase of medical and non-medical goods and services. This included RPR test kits, surgical gloves, BP machines, stethoscope, digital thermometers, urine sticks, suction tubes, cord clamps, and EDTA and BDTA bottles; food and beddings for in-patients; cleaning materials; maintenance of buildings, bicycles, and motorbikes; transport for outreach activities and referrals (taxi and bus fares, fuel, and hire/purchase of new motor bikes and bicycles); office supplies and furniture; kitchen utensils; meetings; construction of maternity shelters; uniforms and dust coats; hiring of additional labour (mostly data clerks but also nurses and midwives in some cases); humpers/motivational packs for traditional birth attendants (TBAs) and mothers; cash for TBAs and other community members; and electricity/energy (solar panels, batteries, inverters, and power generators).

In accordance with the RBF design, the amount which was received by each member of staff was dependent on the individual's performance scores, actual RBF income realized, investment priorities, the number and composition of staff at the health facility, and individual salary levels. As such, RBF staff incentives (bonuses) increased the total income for all health workers at the RBF health facility albeit by different margins/percentages. However, the proportion of RBF staff incentives to government staff salaries was only about 10% during the entire duration of the project (Figure 7).

Figure 7: Use of RBF Funds, and Proportion of RBF staff incentives to Government staff salaries



b) Allocation and use of the RBF matching grant in C1 Health Facilities

The RBF matching grants which were disbursed to C1 districts had some restrictions on their spending namely: a) Resources could only be used for meal allowances or per diems according to the number of days worked, and b) activities had to be related to the delivery of maternal and child health interventions at health facility level. The manner in which health facilities in C1 districts used the RBF matching grant was also dependant on how much was disbursed to the health facility by the DMO. In some C1 districts, part of the RBF matching grants were used for centralized procurements on behalf of the health centres to cover the costs of supervision visits, and the balance distributed to health centres on a per capita basis for expenditure on cleaning materials, stationary, medical, fuel and non-medical items, meal allowances, and incentivizing patients to come to health facilities. This money was treated as imprest, meaning that all health centres in a district had to retire the money

at the DMO before the end of the quarter. This would then allow the DMO to submit a consolidated financial report to the MOH to trigger the release of the next allocation.

Due to these parameters and/or restrictions, most of the health facilities found it difficult to use all the money which was disbursed to them, and in turn, the DMOs also failed to submit consolidated financial reports to the MOH to prompt the replenishment of funds. In other C1 districts, RBF matching grants were not disbursed to health centres at all and the DMOs decided what to buy for the health centres including motor cycles, mass campaigns, meal allowances for district staff and volunteers, health facility maintenance, and medical and non-medical items.

5.3.4 Governance and managerial autonomy at health facilities

55. Use of communities in the management and delivery of health services is one of the key strategies highlighted in Zambia's National Health Strategic Plan (2001-2016) aimed at improving accountability and transparency in resource use and service delivery. This strategy has been in place since 1991/2, and over the years, representative structures (such as Neighbourhood and Health Centre Committees) have been established in communities and are linked to health centres. This sub-section explores whether the RBF strengthened community participation, supervision, performance assessment, and managerial autonomy at health centre levels. Table 21 suggests that health centre committees were more active in RBF facilities (vs. C2) as they reported a significantly higher number of meetings (1.26 more meetings per year on average). And in comparison with both C1 and C2 facilities, RBF facilities also reported more frequent assessment of staff performance, and a higher number of performance assessment at health facilities. District hospitals also conducted more supervisory visits in RBF than C1 facilities.

Table 21: Community participation, supervision, and performance assessment at health facility level

	<i>Mean at baseline</i>			<i>Mean at endline</i>			RBF vs. Control 1		RBF vs. Control 2	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
Facility has a Hospital/Health Center Committee	0.94	0.98	0.98	1.00	1.00	0.96	0.04	0.265	0.08	0.100
Number of members on this Committee	12.40	12.15	12.49	13.27	14.25	13.40	-1.25	0.423	-0.12	0.955
Number of Health Center Committee meetings held in the last 12 months	3.92	4.56	4.77	4.90	4.51	4.53	1.03	0.103	1.26*	0.093
Facility has written records of the Hospital/Health Center Committee meetings	0.71	0.70	0.71	0.87	0.80	0.81	0.04	0.743	0.05	0.674
Facility has a workplan for the current financial year	0.61	0.57	0.53	0.73	0.67	0.67	0.02	0.881	-0.02	0.891
Number of health facility staff meetings held in the last 3 months	2.88	2.82	2.82	2.44	2.50	2.58	-0.19	0.690	-0.29	0.634
Number of visits made by a district hospital representative for supervision	1.06	1.74	1.53	1.42	1.28	1.27	0.82*	0.065	0.66	0.542
Number of visits made by the District Health Management Team for supervision	1.43	1.85	3.60	2.18	2.26	1.84	0.29	0.595	2.49	0.266
Number of visits made by the local government for supervision or technical support	0.19	0.43	0.38	0.50	0.31	0.16	0.41	0.165	0.54**	0.013
Number of times performance of staff assessed internally	2.26	2.31	2.87	5.43	2.13	1.62	3.41***	0.002	4.41***	0.002
Internal staff performance assessment linked to staff salaries or bonuses	0.08	0.14	0.09	0.64	0.18	0.15	0.53***	0.001	0.51***	0.001
Number of times performance of staff assessed externally	1.85	1.96	2.18	3.57	2.26	1.57	1.40**	0.046	2.33***	0.003
External staff performance assessment linked to staff salaries or bonuses	0.09	0.16	0.10	0.43	0.16	0.21	0.34**	0.027	0.25	0.114
Number of times performance of the facility as a whole assessed externally	1.90	1.70	1.84	4.49	3.16	1.80	1.15	0.365	2.64**	0.022

Sample size 348 facilities; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

56. Another design feature of the RBF programme was to introduce more managerial autonomy and devolved decision making at health facilities. These dimensions of management were tracked in the evaluation as facilities were interviewed about the level of autonomy during the follow up survey. The questions were related to the perceived autonomy of the facility in-charge on assigning task to staff, allocating budget, provision of services, and obtaining resources. The responses were recorded on a Likert Scale with values ranging from 1 (least autonomy) to 5 (maximum autonomy). Responses were then converted to a binary scale where higher scores

than neutral (4 and 5) were coded 1 and zero otherwise. An autonomy index was constructed utilizing selected elements of autonomy such as ability to allocate resources and tasks effectively within the facility. The index was further standardized. As shown in Table 22, apart from the overall autonomy index, only two (2) of the individual measures of autonomy were statistically significant. RBF facilities (vs. C2) reported significantly higher autonomy on service provision, clarity on policies and procedures for doing things as well as the overall autonomy index.

Table 22: Managerial autonomy at health facility level

	<i>Mean at baseline</i>			<i>Mean at endline</i>			RBF vs. Control 1		RBF vs. Control 2	
	RBF	Control 1	Control 2	RBF	Control 1	Control 2	Impact estimate	p-value	Impact estimate	p-value
able to allocate my facility budget	0.77	0.68	0.80	0.90	0.80	0.76	0.022	0.873	0.092	0.147
able to assign tasks and activities to staff	0.87	0.86	0.82	0.91	0.91	0.89	-0.001	0.992	-0.014	0.690
DHMT supports my decisions and actions for doing a better job in my facility.	0.87	0.73	0.89	0.85	0.89	0.93	-0.185	0.119	-0.032	0.269
choice over who I allocate for what tasks.	0.77	0.73	0.82	0.77	0.80	0.76	-0.067	0.565	0.035	0.556
choice over what services are provided in the facility.	0.60	0.61	0.76	0.43	0.41	0.33	0.036	0.797	0.127*	0.055
enough authority to obtain the resources I need (drugs, supplies, funding)	0.72	0.80	0.80	0.65	0.60	0.51	0.135	0.500	0.111	0.123
policies and procedures for doing things are clear to me.	0.91	0.98	1.00	0.96	0.93	0.91	0.095	0.180	0.072**	0.016
DHMT provides adequate feedback to me about my job and the performance of my facility.	0.74	0.68	0.67	0.85	0.84	0.82	-0.050	0.719	-0.021	0.635
Autonomy index	-0.04	-0.13	0.19	0.11	-0.03	-0.16	0.06	0.880	0.26**	0.037

Sample size 348 rural health centres; * p<0.1 ** p<0.05 *** p<0.01; impact estimates adjusted for district pair matching with standard errors clustered at district level

5.3.5 Satisfaction and motivation of the health workers

57. This section outlines key findings related to human resources for health, specifically a focus on job satisfaction and motivation of the health workers in RBF facilities compared with control facilities. For each outcome, the impact estimate given is the difference in the change in the outcome for health workers in RBF and control districts between the baseline and endline facility surveys, including facility-level fixed effects as well as linear controls for worker age, gender, and cadre.

58. Job satisfaction is measured using a method based on two existing validated tools, the Minnesota Satisfaction Questionnaire and the Job Satisfaction survey. This contains numerous satisfaction-related questions recorded on a five-point Likert scale. During the analysis, responses to each question were normalized to a 100% scale and then questions were grouped by thematic area, including relationship between staff and outside, working conditions, compensation, recognition and career development. Equal weights were assigned to all

questions within a thematic area. Motivation measures were constructed in a similar way, but with a different set of items, relating to 'intrinsic motivation' and 'extrinsic motivation'. Intrinsic motivation includes 'self-concept', namely an individual's perception of his or her ability to perform, and 'well-being' which is often motivated by financial gain. Extrinsic motivation includes team work, autonomy of staff, working environment, recognition of staff and leadership of facilities.

59. Table 23 shows that the pilot RBF intervention increases the compensation-related component of job satisfaction, implying that workers were more likely to feel adequately remunerated for their work. This was not true for comparisons between C1 and C2. There were also slight gains in satisfaction with work conditions and opportunities when RBF is compared with C2. No other differences were statistically significant. On Table 24, all the results related to worker motivation are statistically insignificant except for well-being where the RBF performed better than C2 facilities.

Table 23: Job satisfaction

	RBF vs. Control 1		RBF vs. Control 2		Control 1 vs. Control 2	
	N	β (s.e.)	N	β (s.e.)	N	β (s.e.)
Relationship within facility	448	-4.155 (2.816)	464	0.477 (1.016)	344	4.941* (2.585)
Relationship outside of facility	448	1.643 (2.964)	464	0.427 (1.491)	344	-0.589 (3.123)
Work conditions	448	6.393 (5.121)	464	4.366* (2.183)	344	2.199 (5.903)
Compensation	448	8.639** (4.081)	464	3.880* (1.994)	344	-0.822 (4.872)
Recognition	448	1.439 (2.842)	464	0.086 (1.324)	344	-1.437 (2.239)
Opportunities	448	4.686 (4.183)	464	3.641* (2.004)	344	2.298 (5.240)

Facility fixed effects adjusted for age, sex, cadre and district pair matching; SEs clustered at district level; * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table 24: Motivation for work

	RBF vs. Control 1		RBF vs. Control 2		Control 1 vs. Control 2	
	N	β (s.e.)	N	β (s.e.)	N	β (s.e.)
Teamwork	447	0.385 (3.132)	462	0.925 (1.429)	345	1.622 (3.511)
Autonomy	448	0.822 (4.311)	463	1.314 (1.768)	345	1.298 (4.488)
Changes in facilities	448	-2.096 (2.664)	463	1.026 (1.240)	345	3.830 (2.640)
Work environment	448	-1.788 (2.597)	463	1.257 (1.260)	345	4.305 (3.028)
Self-concepts	448	-0.727 (1.866)	463	0.774 (1.075)	345	2.214 (2.359)
Recognition	448	-0.380 (3.282)	461	-0.837 (1.330)	345	-0.890 (2.851)
Well-being	450	1.100 (2.981)	465	2.418* (1.236)	345	3.934 (2.499)
Leadership of facility	431	-3.075 (4.885)	446	1.210 (2.613)	339	5.554 (5.151)

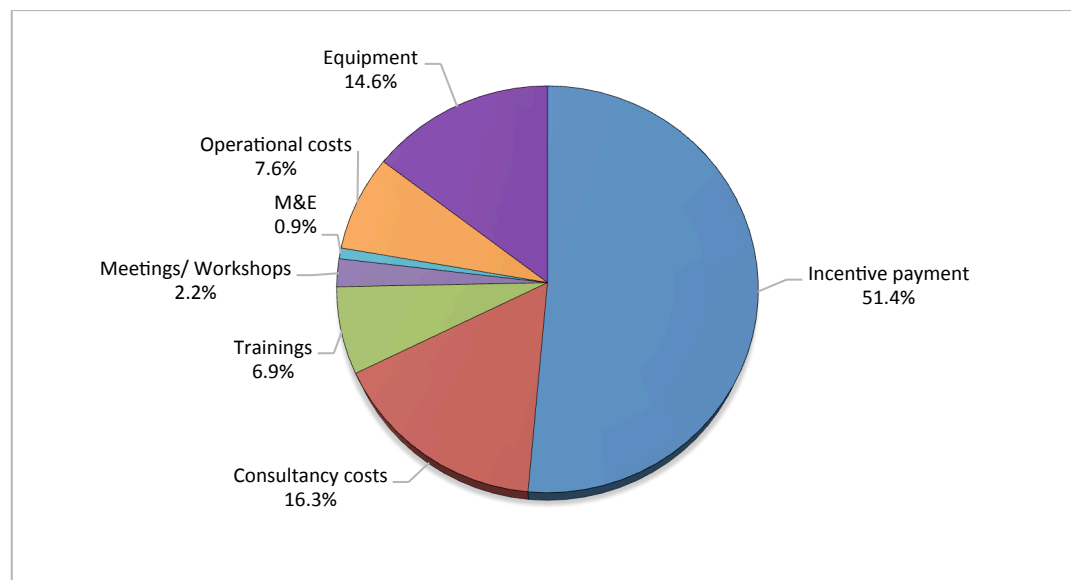
Facility fixed effects adjusted for age, sex, cadre and district pair matching; SEs clustered at district level; * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

5.4 Cost-Effectiveness Analysis

60. The programme costs for the RBF project after 2.25 years is \$13.26 million in total. From this amount, \$10.54 million (\$7.91/capita) was used in the RBF group, while \$2.72 million (\$2.16/capita) was used in the C1 group. The World Bank headquarters programme costs were

\$566,711 out of which \$450,427 (\$0.34/capita) was allocated to the RBF group, and \$116,284 (\$0.09/capita) was allocated to the C1 group. Figure 8 shows the distribution of programme costs for the RBF and C1 groups during the 2.25 years of RBF project implementation.

Figure 8: Distribution of programme costs: RBF and C1 groups combined



Note: M&E denotes monitoring and evaluation

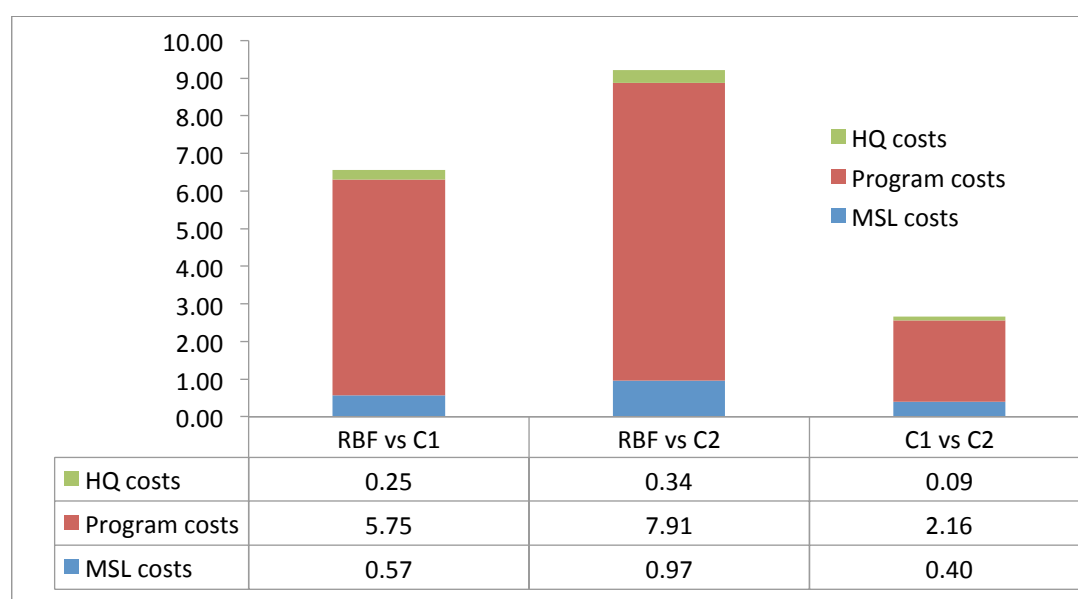
61. For the cost of consumables, Table 25 shows the aggregate expenditure on drugs and supplies from Medical Stores Limited (MSL) before and after RBF for three groups. The drug and supply costs per capita per quarter prior to the RBF programme were estimated as \$0.26, \$0.50 and \$0.42 in the RBF, C1, and C2, respectively. These numbers increased to \$0.59, \$0.77, and \$0.64 during the RBF programme implementation, which shows a slightly higher increase in the RBF group.

Table 25: Consumables expenditures from MSL during the pre- and post-RBF periods in three groups (US\$)

Groups	Pre-RBF period (five quarters)	Post-RBF period (nine quarters)	Pre-RBF period (\$/quarter/capita)	Post-RBF period (\$/quarter/capita)	Difference (\$/quarter/capita)
RBF	1,694,470	6,991,502	0.26	0.59	0.33
Control 1	3,097,135	8,489,457	0.50	0.77	0.27
Control 2	2,924,591	8,062,629	0.42	0.64	0.22

62. Figure 9 shows the incremental (MSL, Programme, and HQ) costs per capita among the three groups during the RBF implementation period (9 quarters). The RBF group cost \$6.56 per capita and \$9.21 per capita more than the C1 and C2 groups, respectively. The C1 group cost \$2.65 per capita more than the C2 group.

Figure 9: Incremental costs per capita over 2.25 years among three groups (US\$)



Note: HQ denotes headquarters; MSL denotes medical stores limited

63. Table 26 shows the coverage and quality of the key maternal health services at baseline and endline from the household and health facility surveys. Comparing RBF with C1, there was a statistically significant difference in Hib vaccination and use of family planning, where the RBF districts had higher utilization than C1 districts. In comparison to C2, the major improvement in the RBF group was in the provision of institutional deliveries (12.2%), PNC (7.8%), IPT (3.0%), vaccinations (BCG, DPT, and HIB) ranging from 5.5% to 19.1%) and family planning (19.5%). However, C1 districts had a significantly greater improvement in institutional deliveries (17.6%), PNC (13.5%), and DPT (6.4%).

64. With regards to quality of key maternal health services, compared to C1 districts, RBF showed a positive impact on the quality of care in institutional delivery (0.7%), vaccination (3.2%) and FP (4.9%). Compared to C2 districts, RBF showed the greatest improvement in the quality of care, by 3.1% for institutional delivery, 2.9% for ANC, 3.8% for vaccination, and 9.7% for family planning. However, C1 districts also had an improvement in quality of care in comparison with C2 districts but at a much lower level of achievement than RBF districts. This is with the exception of quality improvement in PNC where the C1 districts achieved better results against the C2 districts than what the RBF districts achieved against C2 districts.

Table 26: Coverage and quality of key maternal and child health services at baseline and endline

Services	Baseline				Endline				DIDs		
	RBF	C1	C2		RBF	C1	C2		RBF vs C1	RBF vs C2	C1 vs C2
Coverage of key maternal and child services											
Ins Del	68.3%	56.4%	70.9%		80.8%	74.3%	71.2%		-5.4%	12.2%**	17.6%***
ANC	97.5%	96.2%	96.3%		98.9%	99.0%	99.1%		-1.4%	-1.4%	0.0%
PNC	70.3%	56.0%	76.4%		82.4%	73.8%	80.7%		-5.7%	7.8%*	13.5%***
BCG	95.6%	97.8%	97.6%		100.0%	99.5%	95.6%		2.7%	6.4%*	3.7%*
DPT	97.1%	95.2%	95.8%		98.6%	97.6%	91.8%		-0.9%	5.5%*	6.4%*
HIB	82.5%	88.3%	81.8%		97.9%	88.7%	78.1%		15.0%***	19.1%***	4.1%
IPT	92.0%	92.4%	95.1%		98.0%	96.1%	98.1%		2.3%	3.0%**	0.7%
FP ^A	6.5%	9.9%	7.7%		34.0%	15.6%	15.7%		21.8%**	19.5%**	-2.3%
Quality index of key maternal and child services [†]											
Ins Del	65.5%	66.8%	67.0%		73.5%	74.1%	71.9%		0.7%	3.1%	2.4%
ANC	66.9%	69.1%	68.6%		75.0%	77.2%	73.8%		0.0%	2.9%	2.8%
PNC	66.7%	68.4%	68.3%		74.1%	76.6%	73.4%		-0.8%	2.3%	3.0%
Vaccination	78.7%	80.7%	81.7%		81.2%	80.0%	80.4%		3.2%	3.8%	0.6%
FP	77.7%	78.6%	80.6%		81.6%	77.6%	74.8%		4.9%	9.7%	4.8%

Note: Ins Del denotes institutional deliveries. ANC denotes any antenatal care, and was included in the analysis as it is an important mother and child health service; PNC denotes any postnatal care; BCG denotes Bacillus Calmette–Guérin vaccine; DPT denotes diphtheria, pertussis and tetanus vaccine; Hib denotes Haemophilus influenza type b vaccine, IPT denotes intermittent preventive treatment; FP denotes family planning; DID denotes difference in differences; ^aData are from the health facility survey. *p<0.10, **p<0.05, ***p<0.01. Linear probability model with difference-in-difference specification, including controls for district stratification. Errors are clustered at the district level. [‡]Quality index was constructed for each arm through averages of key quality dimensions across the health facilities, and thus no statistical tests were conducted.

65. Over the RBF implementation period (2.25 years) and among the total 1.3 million population, without including quality of care in the analysis, the RBF programme resulted in saving 11 lives for pregnant women, and 214 lives for children under five, compared to the C1 group, and saving 22 lives for pregnant women, and 497 lives for children under five, compared to the C2 group (Table 27). The number of lives saved increased after including the improvement of quality of care in the analysis. Specifically, compared to the C1 group, the RBF programme saved 279 lives for mothers and children (with lower and upper bounds of 214 and 324, respectively). And in comparison to the C2 group, RBF saved 641 lives (with lower and upper bounds of 580 and 700, respectively).

Table 27: Lives saved from the RBF programme in comparison with lower and upper bounds *

	Number of deaths obtained from LiST					Lives saved (quality unadjusted)			Lives saved (quality adjusted)		
Population	RBF	C1	C2	C1 quality adjusted	C2 quality adjusted	RBF vs C1	RBF vs C2	C1 vs C2	RBF vs C1	RBF vs C2	C1 vs C2
Children < 5											
2013	4 478	4 537	4 636	4 553	4 673	59	158	99	75	195	120
2014	4 334	4 489	4 673	4 524	4 752	155	339	184	190	418	228
Subtotal	8 812	9 026	9 309	9 077	9 425	214	497	283	265	613	348
Maternal deaths											
2013	141	145	149	146	151	4	8	4	5	10	5
2014	133	140	147	142	151	7	14	7	9	18	9
Subtotal	274	285	296	288	302	11	22	11	14	28	14
Total lives saved (point estimate)						225	519	294	279	641	362
Lower bound						167	461	226	214	580	293
Upper bound						267	576	356	324	700	430

*Negative value indicates the dominant effect of control group 1 over RBF group. C1 denotes the control group 1 and C2 denotes the control group 2. LiST denotes Lives Saved Tool.

66. The number of lives saved were later expressed in QALYs (Table 28). If the analysis is not adjusted for quality, the RBF programme gained 5,325 QALYs as compared to C1; and 12,291 QALYs as compared to C2. When quality of care is included in the analysis, the RBF programme gained 6,602 QALYs as compared to C1; and 15,178 QALYs as compared to C2. (Table 28). For the C1 vs C2 analysis, the C1 group gained 6,966 QALYs and 8,576 QALYs if the analysis is unadjusted and adjusted for quality, respectively.

Table 28: QALYs gained from the RBF programme in comparison with controls, with lower and upper bounds*

Population	RBF vs C1		RBF vs C2		C1 vs C2	
	Mid-point (lower bound; upper bound)		Mid-point (lower bound; upper bound)		Mid-point (lower bound; upper bound)	
	QALYs gained (unadjusted for quality)	QALYs gained (adjusted for quality)	QALYs gained (unadjusted for quality)	QALYs gained (adjusted for quality)	QALYs gained (unadjusted for quality)	QALYs gained (adjusted for quality)
Pregnant women	237 (216-302)	302 (237-345)	475(425-539)	604(539-626)	237(176-302)	302(237-345)
Children under 5	5,088(3,733-6,015)	6,300(4,826-7,323)	11,816(10,480-13,100)	14,574(13,195-15,953)	6,728(5,171-8,131)	8,274(6,704-9,843)
Total	5,325 (3,948-6,317)	6,602(5,064-7,688)	12,291(10,905-13,639)	15,178(13,734-16,579)	6,966(5,347-8,433)	8,576(6,942-10,188)

*Negative values indicate the dominant effects of control group 1 over the RBF group. QALY denotes quality adjusted life year

67. Table 29 shows the incremental cost-effectiveness ratios (ICERs) of the RBF programme, in comparison with the two control groups. It shows that compared to the two controls, the RBF programme was cost-effective. The ICERs were \$1,642 per QALY gained and \$999 per QALY gained, when compared with C1 and C2, respectively, without quality adjustment. These ratios improve to \$1,324 per QALY gained and \$809 per QALY gained, when compared with C1 and C2, respectively, if quality of care is added. In 2013, (mid-point of the RBF project implementation in

Zambia), the gross domestic product (GDP) per capita was \$1,759 (World Bank, 2015). Since the ICERs are less than the GDP per capita for Zambia in 2013, the RBF programme is cost-effective in comparison to C1 and C2. However, C1 is also cost-effective as compared to C2. The ICERs for C1 vs C2 were \$508 and \$413 per QALY gained, without and with quality adjustment, respectively.

Table 29: Incremental cost effectiveness ratios, with lower and upper bounds*

Comparison	Cost/life saved (US\$)	Cost/QALY gained (US\$)
	Mid-point (lower bound; upper bound)	Mid-point (lower bound; upper bound)
RBF vs C1 (unadjusted)	38,857 (32,744-52,351)	1,642 (1,384-2,214)
RBF vs C1 (quality adjusted)	31,336 (26,983- 40,853)	1,324 (1,141-1,727)
RBF vs C2 (unadjusted)	23,666 (21,324-26,643)	999 (900-1,126)
RBF vs C2 (quality adjusted)	19,161 (17,546-21,177)	809 (741-895)
C1 vs C2 (unadjusted)	12,040(9,943-15,663)	508 (419-662)
C1 vs C2 (quality adjusted)	9,999 (8,232- 12,081)	413 (348-510)

*Negative values indicate the dominant effect of control group 1 over the RBF group. C1 denotes the control group 1 and C2 denotes the control group 2. QALY denotes quality adjusted life year.

68. In sum, the cost-effectiveness analysis estimates that RBF delivered greater health gains, in terms of lives saved or QALYs gained, than C1 and C2. However, these gains were supplied at a higher unit cost. Furthermore, both RBF and C1 were found to be cost-effective when the ICERs are compared with the GDP per-capita of \$1,759 in 2013 in Zambia.

6. Discussion and Conclusion

69. The Zambia RBF pilot programme was designed to strengthen the health system and improve the coverage and quality of MCH related health services. A prospective quantitative impact evaluation assessed the effectiveness, and cost-effectiveness of the RBF pilot. This section summarizes the key results and lessons learnt from this exercise. Two additional policy questions – on the level of incentives and the likelihood of audit, which were also part of the overall evaluation, are discussed in detail in the appendices.
70. The RBF was designed to have a positive effect on the quantity and quality of targeted MCH services, and functionality of the health system. The evaluation investigated the impact over a broad range of targeted and non-targeted indicators related to maternal and child health services. Of the nine indicators directly targeted by the RBF programme through the incentive structure, seven were directly measured or proxied in the population.¹³ Some of the measures responded to the RBF intervention, with a broadly similar set also showing improvements under the enhanced financing arm (C1). Most notably, institutional deliveries (in-facility delivery rate) in RBF districts increased by approximately 13 percentage points relative to the pure control districts (C2). The same indicator rose by 18 percentage points in C1 districts relative to C2, suggesting larger gains for this indicator in the enhanced/input financing arm. Results for deliveries by skilled providers also show improvements in both the RBF and C1 districts relative to C2, but the C1 arm had greater magnitude. The higher magnitude in C1 districts for institutional deliveries and skilled birth attendance suggests relative effectiveness of C1 districts as compared with the RBF programme for these two measures.
71. We further observe that neither RBF nor C1 districts experienced gains in the targeted ANC indicator, perhaps due to the already high rates of ANC coverage in the population. However, a few non-directly targeted coverage indicators saw improvements under the RBF particularly the timing of first ANC. The timing of the first ANC visit improved by about two weeks in the RBF arm as compared to both controls. This is an important gain in maternal care that is seldom observed in a broad-based primary care intervention such as RBF. While there were broad-based gains in the RBF and C1 arms in postnatal coverage (PNC), the rate of change for any PNC was more rapid in C1 districts as compared to RBF districts. However, the results show no differences in the relative gains in the provision of PNC at a facility by a qualified provider for both RBF and C1 districts in comparison to C2 districts.
72. For family planning (FP) services, the results show that the use of modern or any contraceptive method remained relatively unchanged at endline as compared to the baseline across the 3 study arms. But further review of the availability of FP outreach services shows a rise in the provision of any FP outreach programme in the RBF districts vis-à-vis the C2 but a greater rise in FP outreach in C1 districts when compared with both RBF and C2 districts. For immunization services, we observe that immunization coverage in the surveyed areas of rural Zambia was

¹³One of the remaining indicators involved special sub-population of HIV positive pregnant women whose coverage was not tracked by the data collection. The other remaining indicator applies to all women of reproductive age and not just those with a recent birth and hence were not sufficiently represented in the household survey.

erratic during the RBF implementation period. In the household survey data, rates of full vaccination declined in both C1 and C2 districts but remained constant or slightly higher in RBF districts. This suggests that the RBF programme was protective with respect to some measures of immunization coverage – any immunization and DPT injection – which were significantly higher than in RBF communities as compared to C2. However, these results were not precisely estimated.

73. As regards to structural quality, results on the RBF vs C1 were largely inconclusive but the RBF districts performed better than C2 districts in terms of the status of infrastructure and availability of functional medical equipment. The quality of delivery rooms in RBF facilities was better than the delivery rooms in C1 and C2 districts while the quality of curative care in RBF facilities was better than C2 facilities.
74. Process quality during maternal and child health care was not directly targeted by the RBF programme (with the exception of the two process measures tied to ANC – IPT and HIV testing). The evaluation measured mother's knowledge of danger signs during pregnancy which showed that women residing in RBF districts are significantly more likely to list several out of the 12 danger signs as compared to those residing in C1 districts who were not able to list any.
75. Despite higher knowledge, results from the household survey showed minimal progress on process quality of maternal health care under the RBF programme except for the provision of a tetanus injection (vs C1); any iron tablets and malaria drugs were higher in RBF communities than C2 communities. C1 districts witnessed better improvements in blood tests and any iron taken during ANC than the RBF. Results from patient recall showed that more women reported to have received advice on diet in RBF facilities (vs C1), and having had their abdomen measured and palpated (vs C2). However, women who attended C1 facilities reported to have received explanations on the side effects of iron folic acid tablets as compared to those who went to RBF facilities. The results also showed no gains in process quality for postnatal care in RBF communities. On the other hand, mothers from C1 communities reported higher immediate initiation of breastfeeding and receipt of Vitamin A after delivery as compared to both RBF and C2 communities.
76. Clients who visited RBF health facilities were more satisfied with the time that the health workers spent with them. The data shows that health workers in RBF facilities spent sufficiently more time during consultations with their patients as compared to both C1 and C2 health facilities. There was also more trust in health workers operating in RBF facilities as compared with C1 facilities for both maternal and child health services.
77. When it comes to understanding the causal and behavioural mechanisms through which RBF and enhanced financing (C1) achieved these gains, the evaluation partially investigates this question. The health worker interview found that the level of job satisfaction of health workers increased as a result of the RBF and health worker turnover was lower, suggesting that more engaged health workers with more experience in the catchment area played a role. These gains in satisfaction and retention are relatively larger in RBF areas than in C1 areas indicating the likely influence of staff incentive payments (which were not present in C1).

78. When investigating the role of staff incentives in determining RBF effectiveness, the power of the incentive is a critical aspect to note—individuals in general exhibit a greater response to higher monetary incentives. In terms of the relative power of the RBF incentive, the amount which was received by each member of staff was dependent on the individual's performance scores, actual RBF income realized, investment priorities, the number and composition of staff at the health facility, and individual salary levels. Consequently, health workers received about 10% of their official staff salaries on average as RBF staff incentives. At the start of the RBF project, the proportion were higher but 6 months after the start the RBF project, the Zambian Government increased staff salaries for all civil servants ranging from 100% to 200%. While there is little empirical evidence on what constitutes optimal incentive levels – either at the facility or individual worker level – to foster maximal effectiveness of an RBF-type mechanism, evidence shows that small incentives often result in no appreciable gain in targeted outcomes (Friedman and Scheffler, 2015). The relatively small proportion of total health worker remuneration coming from the RBF mechanisms suggests that greater gains may have been possible if the RBF incentives were higher.
79. Determinants of programme effectiveness also include contextual and implementation factors. Some of these relate directly to the power of the individual worker incentive. While the above staff incentive arrangement was designed to facilitate an increase in staff incomes, several health facilities agreed with their staff members to give up whole or part of their incentive/bonus during a particular quarter in order to make a large investment, e.g. purchase of a motor cycle, water pump, etc. This could be considered a sign of dedication to improving the welfare of the community, and/or altruism. But in most cases, such capital investments could be spread across a number of quarters which can affect staff motivation.
80. In terms of non-wage resources, RBF performance grants at the health facility level complemented GRZ resources significantly. The results show that the total RBF performance grant was about 6 times the value of the GRZ grant over the project period.¹⁴ However, the RBF grant was growing faster than the GRZ grant and the latter actually declined between 2013 and 2014. There could be several explanations for this but a study by Dusseljee *et al.* (2014) observed that the district management in the RBF health facilities were reducing the amount of the GRZ grant that was being disbursed to the health facilities. This suggests that there may have been aid fungibility¹⁵ or substitution of financing because the proportion of the GRZ grant to the RBF grant decreased by half between 2012 and 2014. This further suggests that the RBF grant was not additional to the existing financial resources at the RBF health facilities in accordance with the project objectives. This has a number of policy implications on aid effectiveness as a whole, and efficacy of the RBF programme. The RBF funds may just have substituted rather than complemented GRZ spending. To mitigate this problem, future RBF programs could consider putting in place indicators linked to GRZ budget performance at national and district levels to ensure that the RBF grants are additional to GRZ grants.

¹⁴ Over the project period (2012-2014), the total GRZ operational grant was only 18% the total value of the RBF performance grant

¹⁵ Aid fungibility is when donor funding for health substitutes for—rather than complements—health financing by recipient governments.

81. Given the high levels of RBF grant funding, far above the GRZ grant, questions may be raised on the future financial sustainability of the RBF programme. However, considering that only half of the RBF funds were being used for operational activities while the rest were spent on staff incentives, sustainability may not be an issue. The Zambian Government can easily absorb this funding while the loss in staff incentives might not have a huge impact since the proportion of the staff incentives to the staff salaries was only 10%. It should also be noted that the GRZ was responsible for staff salaries which were far much higher than the RBF grants.
82. Apart from financial sustainability, the study demonstrates that RBF can be successfully implemented through a “contracting-in” public health system using the existing government systems and structures in Zambia. In the long run, this approach could facilitate institutional and impact sustainability. This is highlighted in the Paris Declaration on Aid Effectiveness, as well as other studies on aid effectiveness^{16,17} where the common agreement is that using a country’s own institutions and systems to implement projects can strengthen a country’s capacity to implement programmes, and programmes being implemented can be sustained.
83. In contrast with RBF facilities, health facilities in C1 districts could not spend the matching grant on staff incentives which were about 47% on average in the RBF districts. The money disbursed to the health facilities in C1 districts was required to be retired before replenishment which caused further disbursement delays. This was contrary to health facilities in the RBF arm where RBF payments were disbursed directly in the health centre bank accounts’ and didn’t needed to be retired. Additionally, in terms of autonomy over the use of funds, RBF health facilities were undoubtedly better than C1 health facilities. The results show that the funds for C1 health facilities were not being fully disbursed from the C1 district to the C1 health facilities. Instead, managers in several C1 districts used part of this money (which was solely meant for C1 health facilities) for centralized procurements, and only disbursed the balance of what remained. A study by Dusseljee et al. (2014) confirms this finding. On the contrary, in the RBF arm, the intended quantity of money reached the health facilities because it was sent directly into the health facility bank accounts. This facilitated fiscal decentralization and greater autonomy over resources at facility level.
84. To contextualise the discussion on managerial autonomy at health facility level, the study shows that RBF facilities (vs. C2) reported significantly higher autonomy on service provision, clarity on policies and procedures for doing things as well as the overall autonomy index. The study further reveals that RBF health facilities reported more frequent assessment of staff performance, and a higher number of performance assessment at health facilities in comparison with both C1 and C2 facilities. District hospitals also conducted more supervisory visits in RBF than C1 facilities. In addition, health centre committees were more active in RBF facilities (vs. C2). These findings demonstrate greater accountability and transparency in planning, resource use, service delivery, and community participation. In Appendix 2, the results also show that RBF was successful in improving the accuracy of reporting for some indicators (deliveries and PMTCT) as compared to C1 and C2.

¹⁶ <http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=141&Lang=en>

¹⁷ Institute for Health Sector Development (2004) cited by Vergeer and Chansa (2008).

85. The relevant point from the design perspective is that disbursement of RBF financing directly to health facilities facilitated fiscal decentralisation. The study was designed to equalize total RBF financing between health facilities in RBF districts and C1 districts. By using two different disbursement mechanisms, the study was able to measure the success of each system in terms of overall level of RBF funding being utilized.¹⁸ Results show that health facilities in the C1 districts did not receive the same amounts as the RBF districts due to delayed retirement and low absorptive capacity. By the end of the RBF programme, the proportion of disbursement to C1 districts was only 56% of what the RBF districts had received. Disbursements to C1 districts lagged behind mainly due to: i) delays in disbursing the funds from the district accounts to health facilities as an imprest, and ii) delayed retirement by health facilities which in turn contributed to delayed replenishment of the district account.¹⁹ It is clear that disbursement mechanisms affect both absorptive capacity and the level of available funding.
86. The study was able to explore some of the causal and behavioural mechanisms through which the RBF and C1 could have achieved and/or not achieved gains in the targeted indicators. For the enhanced financing (C1) arm, the key question is whether the gains were the result of availability of inputs, increased financing, earmarking of funds for priority maternal and child health interventions, or other factors. A corollary question is whether greater gains could have been observed in the C1 arm if financial flows to C1 facilities actually equalled those received by RBF facilities. As earlier stated, the IE had 3 districts in each province (RBF, C1, C2) and these districts were being managed by the same Provincial Medical Office (PMO). In line with Government guidelines, all districts in a province attend quarterly GRZ implementation review meetings and it is possible that during these meetings there could have been cross-pollination of ideas. Consequently, health facilities in the C1 districts may also have been implementing RBF initiatives and could have behaved as if they were incentivized. For example, one World Bank supervision mission noted that some C1 districts were using some form of output-based approaches. With no concealed investigation, the study units were aware of the experiment and the C1 districts could have tried to out-perform the RBF districts.
87. For the RBF, the key question is whether the RBF districts could have achieved even more. In exploring this question, we noted that the Zambia RBF project was being implemented in a health system that already had high coverage in some of the key MNCH indicators being incentivized. As such, perhaps it would have been more prudent to have implemented a target or coverage based performance incentive framework rather than fee-for-service. Furthermore, the results show that health workers received about 10% of their official GRZ staff salaries on average as RBF staff incentives by the end of the pilot period. This may not have been sufficient to have induced change as discussed above.

¹⁸ RBF performance grants were being disbursed directly into bank accounts at RBF health facilities while the matching grants for health facilities in C1 districts were being disbursed through bank accounts at District level

¹⁹ Funds disbursed to C1 health facilities needed to be retired (accounted for at central level through proof of receipts and other supporting documents) before replenishment

88. The CEA shows that the RBF (vs C1) provided more total health benefits (QALYs gained) but at a higher unit price. Nonetheless, in comparison with the two control groups, the RBF programme is a cost-effective approach in improving maternal and child health. When the RBF group is compared with the C1 group, the mid-point ICER is \$1,642 per QALYs gained (without quality adjustment), and \$1,324 per QALYs gained (with quality adjustment). When the RBF group is compared with the C2 group, the mid-point ICER is \$999 per QALYs gained (without quality adjustment), and \$809 per QALYs gained (with quality adjustment). All these values are less than the GDP/capita of \$1,759 in 2013 (mid-year of RBF programme) in Zambia^{20,21}. Since these ICERs are less than Zambia's GDP per capita in 2013, the RBF programme was cost-effective in comparison to C1 and C2. However, the input financing approach (C1) was also cost-effective in comparison to C2. The ICERs for C1 vs C2 were \$508 and \$413 per QALY gained, without and with quality adjustment, respectively. Thus, depending on which group is used for comparison, the ICER varies, but the estimates point to a cost-effective impact of both the RBF and C1 groups than the C2 group.
89. For the CEA analysis, it should be observed that health system investments and gains that may have occurred only in the RBF group weren't fully evaluated. In addition, the confidence bounds around these estimates are not able to definitively distinguish the two approaches partially due to the uncertainty inherent in CEA studies.²² Nonetheless, we can conclude that both the RBF and C1 are cost-effective when compared with Zambia's level of income in 2013.²³
90. The overarching conclusion is that both the RBF and C1 contributed to increased utilisation of key MNCH services in Zambia. However, as compared to the C1, RBF had a more positive effect on health systems governance particularly availability of equipment, structural quality, managerial autonomy, accuracy in reporting, satisfaction and retention of health workers, and level and predictability of funding. Internal and external verification of results, and regular supportive supervision which were a key feature in the RBF districts could have contributed to these successes. Another feature is that purchasing mechanisms were enhanced in the RBF and this potentially contributed to greater efficiency and value for money. These important elements could not be achieved in the input financing (C1) arm.

²⁰ The World Bank. GDP per capita (current US\$). Washington, DC: The World Bank; 2015 [cited 2015 Sept 30]; Available from: <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>.

²¹ WHO recommends comparing ICER to GDP/Capita. GDP/Capita proxies for the productivity a person in a year. If an intervention could save more than what a person produces in a year, it is regarded as highly cost-effective.

²² The use of expert opinions to estimate the health gains from increases in quality of care is sensitive to an expert's understanding of the exercise, their knowledge of the subject matter, and the disconnection between the concepts (used in Delphi consultation) and actual measures (obtained from survey).

²³ WHO recommends comparing ICER to GDP/Capita. GDP/Capita proxies for the productivity a person in a year. If an intervention could save more than what a person produces in a year, it is regarded as highly cost-effective. Overall, RBF programme is very cost-effective, whether it is compared to controls 1 or 2 group.

6.1 Lessons Learnt

91. The results from the study shed light on several areas which have been under discussion as regards to the RBF in terms of results, implementation and evaluation processes:

- (i) The study demonstrates that an RBF programme can be successfully implemented to increase delivery of key health indicators through “contracting-in” a capacity constrained public health system. Many other examples of successful public sector RBF programs occur in middle-income countries (i.e. Argentina) or when implemented by a specialist third party (i.e. Zimbabwe). Since Zambia implemented the RBF by using existing government systems, structures, and local expertise, it is potentially easier to scale-up a countrywide RBF programme. This is because the Zambia RBF design allows for financial, institutional, and impact sustainability.
- (ii) It is important to have a routine process evaluation (PE) system in place to continuously monitor the results and overall implementation of the RBF programme. The Zambia RBF pilot programme benefited from a PE system which provided regular updates and insights on the implementation to allow for midcourse corrections and evidence-based policy and planning.
- (iii) While the “contracting-in” design could be potentially more institutionally sustainable, consistency in leadership is a critical component to moving from a project to a programmatic approach that is fully embedded in the larger health sector. In the case of Zambia, there were several exogenous shocks (governance issues, split of ministries, high staff turnover at all levels of Government etc) which made it difficult to have continuous policy dialogue on RBF. To help ensure integration of experiences and lessons of current and future RBF programs into broader health sector dialogue, these programs should be firmly embedded in the planning department of the Ministry of Health (MOH) with co-leadership with a relevant technical unit such as Mother and Child Health. The implementation structure could consist of a mix of dedicated technical civil service staff. In addition, an RBF coordination committee governed by the MOH should bring together interested donors together with Government to discuss emerging results, policy impacts, and the way forward.
- (iv) The Zambia RBF project demonstrates that having a performance incentive framework (provider payment mechanism) linked to targets and production capacity instead of a payment mechanism for all services rendered is potentially better. The Zambia RBF was implemented in a health system that already had relatively high coverage in some of the key maternal and child health indicators. As such, rather than fee-for-service, it may have been more effective to have used a target or coverage based performance incentive framework.
- (v) The enhanced financing arm as part of the evaluation is critical in order to be certain that effects in the RBF arm are not only due to additional financial resources. As shown in the evaluation results, enhanced financing, can also produce good results. In Zambia, these results go a step further in highlighting a potential issue in the current health system

related to funding constraints and unpredictability. Input financing with parameters focused on key interventions can work—but in the case of Zambia, there were also issues in utilizing funds in Control 1 districts which points to disbursement mechanism issues which were not experienced when disbursing RBF grants in the RBF arm. Thus, effective approaches, including direct disbursement of funds to front-line service delivery levels coupled with a variety of financing mechanisms can have a positive impact on service delivery through improved budget performance (disbursement and absorption of funds).

- (vi) Direct disbursement of funds to front-line service delivery levels and use of an effective disbursement mechanism can also increase predictability of funding and managerial autonomy. However, the RBF funds may have substituted rather than complemented government funding due to the poor disbursement of Government grants to pilot health facilities by the district management during the implementation of the RBF project. To mitigate this problem, future RBF programs in Zambia could consider putting in place indicators linked to government budget performance at national, provincial, and district levels to ensure that RBF grants at service delivery levels are additional to government grants.
- (vii) Adequate levels of incentives need be offered to providers to trigger sufficient behavioural change. The relatively low power of RBF staff incentives in relation to guaranteed individual staff salaries (which declined to 10% over the project period) may have limited some of the possible achievable gains by RBF. Furthermore, higher incentive payments in remote areas did not result in increased health outcomes either (Appendix 1). This suggests that provider effort may be relatively inelastic at small incentive levels. This probably explains why the RBF had no impact on the motivation of health workers but had a positive impact on health worker satisfaction, reduced attrition, and responsiveness to the client. Given the high cost of living in Zambia, the additional income from the RBF staff incentive could have been inadequate to fully influence personal behaviour. Future RBF programs should provide adequate but sustainable levels of RBF staff incentives.
- (viii) When introducing the concept of data verification in a health system with little previous experience, repeated outreach to facility management combined with experiential learning may be necessary for management to internalize the reality of a verification audit. This also applies to the possible ramifications for mis-reporting. The audit experiment discussed in Appendix 2 demonstrates a very low level of understanding of the audit likelihoods despite repeated announcement to the facility management, as well as a lack of understanding over mis-reporting thresholds and possible sanctions. As such, the audit likelihood experiment largely failed as the reporting principals were unaware of the likelihood assigned to the facility. Nevertheless, despite discrepancies in reporting found in RBF facilities by the external verifiers, these discrepancies appear to be within the bounds of normal reporting error as they are not significantly different from a sample of C2 facilities.
- (ix) A key component of the Zambia HRBF IE is the cost-effectiveness analysis (CEA) which justified the value of the RBF on both the costs and effectiveness (by increasing both quality and quantity of services). By adding a complementary cost-effectiveness study, the Zambia HRBF IE showed that a number of decisions must be made in the health facilities

on health systems inputs such as personnel, drugs, equipment, buildings, verification, supportive supervision and so forth. The existence of both fixed and variable costs are important aspects in evaluating how much it costs to implement a RBF programme, and the efficacy of RBF programs as compared to non-RBF programs.

- (x) To our knowledge, this CEA study is among the few to incorporate the quality of care in the cost-effectiveness modelling, and the study innovatively uses a Delphi panel to generate a quality index from household survey based results and to convert a quality index to a health effect index. Given that improving the quality of care is one of the major components of the RBF programme, RBF evaluation models should always include an assessment of quality improvements to fully estimate the cost-effectiveness of the RBF programme.

7. Limitations of the study

- 92. Due to budgetary limitations and the high cost of primary data collection in Zambia, population based data was only collected in 18 of the 30 study districts, leading to the possible influence of potential unobserved confounders at the district level for the estimated population level impacts. This limitation is discussed in the Methods Section and alternative p-values, based on randomization inference, are presented. Appendix 4 contrasts these Fisher exact p-values with the more usual (but inapplicable) asymptotic p-values clustered at the survey enumeration area level.
- 93. Secondly, contrary to study design, the amount of funds available to C1 facilities did not equal, as anticipated, to the mean amount earned by an RBF facility due to differences in disbursement mechanisms.
- 94. Lastly, RBF is a comprehensive intervention package including devolved autonomy and enhanced monitoring, supervision, and data verification. The evaluation design was not able to investigate the relative effectiveness of each of these RBF components on the priority outcomes but rather the summary effect of all.

References

- Alonge, O., Gupta, S., Engineer, C., Salehi, A. S., & Peters, D. H. (2014). Assessing the pro-poor effect of different contracting schemes for health services on health facilities in rural Afghanistan. *Health policy and planning*, czu127.
- Avenir Health. (2015). *Spectrum Manual: Spectrum System of Policy Models*. Glastonbury, CT: Avenir Health.
- Basinga P., Gertler P.J., Binagwaho A., Soucat A.L., Sturdy J., Vermeersch C.M. (2011). Effect on Maternal and Child Health Services in Rwanda of payment to Primary Health Care providers for performance: An Impact Evaluation. *Lancet* 377 (9775):1421-8.
- Boschi-Pinto C., Young M., Black R.E. (2010). The Child Health Epidemiology Reference Group reviews of the effectiveness of interventions to reduce maternal, neonatal and child mortality. *International Journal of Epidemiology*. 39 Supplementary 1:i3-6.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3), 414-427.
- CSO (Central Statistical Office), Ministry of Health, Tropical Diseases Research Centre, University Teaching Hospital, University of Zambia, ICF International. (2014). *Zambia Demographic and Health Survey 2013-14*. Lusaka: CSO
- Dusseljee J., Chishimba P., Phiri M. (2014). *Technical Review of the Pilot Results Based Financing (RBF) Project*. Lusaka: Ministry of Health
- Ernst, M. D. (2004). Permutation methods: a basis for exact inference. *Statistical Science*, 19(4), 676-685.
- Friedman, J., and R. Scheffler. (2015). "Pay for Performance in Health Systems: Theory, Evidence, and Case Studies", in *The Global Health Handbook*, edited by Richard Scheffler, World Scientific Press.
- Friedman J., Mutasa R., Mafaune P., Nyameru R., Das A. Forthcoming. "The Impact of Zimbabwe's Health Results Based Financing National Pilot on the Utilization and Quality of Care". World Bank Policy Research Working Paper.
- Huillery, E., & Seban, J. (2014). *Pay-for-Performance, Motivation and Final Output in the Health Sector: Experimental Evidence from the Democratic Republic of Congo*. Working Paper, Department of Economics, Sciences Po, Paris.
- Ministry of Health (2011). *Operational Implementation Manual for Results Based Financing (RBF) in Pilot Districts in Zambia*. Lusaka: Ministry of Health
- Ministry of Health and Child Care (2014). *Zimbabwe National Results Based Financing Approach: Programme Implementation Manual*. Harare: Ministry of Health and Child Care
- Sassi F. (2006). Calculating QALYs, Comparing QALY and DALY calculations. *Health Policy and Planning*. 21(5):402-8.
- Singh S., Darroch J.E., Ashford L.S. (2014). *Adding It Up: The Costs and Benefits of Investing in Sexual and Reproductive Health*. New York, NY: Guttmacher Institute

Stenberg K., Axelson H., Sheehan P., Anderson I., Gulmezoglu A.M., Temmerman M., et al. (2014). Advancing social and economic development by investing in women's and children's health: a new Global Investment Framework. *Lancet*. 12;383(9925):1333-54.

The World Bank (2015). GDP per capita (current US\$). Washington, DC: The World Bank. Retrieved on September 30, 2015 from: <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>.

Vergeer P., Chansa C (2008). *Payment for Performance (P4P) Evaluation: Zambia Country Report for Cordaid*. Amsterdam: KIT Development Policy & Practice

World Health Organization. (2015a). *Zambia: Global Health Observatory Data Repository*. Geneva, Switzerland: WHO. Retrieved on September 30, 2015 from: <http://apps.who.int/gho/data/view.main.61850?lang=en>.

World Health Organization. (2015b). *Zambia: WHO statistical profile*. Geneva: WHO

Appendix 1

Research Question 2 – Do higher incentive payments in remote areas result in increased health outcomes and greater retention of staff?

Appendix 1 evaluates the effect of higher RBF incentive payments in remote health facilities on maternal and child health output and quality measures, health system, and functionality of health workers. Data comes from the two rounds of the health facility survey conducted in RBF and control facilities at baseline and endline. The impact estimate given is the relative change in indicators for the RBF facilities compared between health facilities where the prices of indicators were pegged at 25% more than the normal prices for the RBF indicators (henceforth written as enhanced facilities) versus health facilities that were earning normal prices on each RBF indicator (written as standard facilities). This is the same difference-in-difference framework used in the main report.

A1.1 Quality of services

A1.1.1 Structural quality

This section evaluates the effect of the RBF intervention on facility infrastructure and availability of essential drugs and equipment. The status of infrastructure at the facilities was assessed with direct observation. Relevant dimensions of infrastructure were availability of power, water, telecommunication systems, disinfectants, an outpatient consultation room, availability of key elements in the outpatient room for optimal service delivery, and provision of biomedical waste disposal. An infrastructure index was constructed, including the following items with equal weight: continuous availability of power, water, communication and disinfectants, provision of sharp disposal and basin with soap and water in outpatient room. As shown in Table A1, there were no significant differences between elevated and standard facilities for all indicators except for disinfectant stock-out that was significantly higher in elevated facilities (33% points at 10% significance level).

Facilities were asked if they had availability of specific drugs on the day of survey and for the previous 30 days. Drugs included general antibiotics, analgesics, family planning, anti-malarials, anti-tuberculosis, antiretroviral, emergency obstetric care (EMOC), vaccine, diagnostic kits, fluids and electrolytes. A drug availability index was constructed assigning equal weight to the individual items and was further standardized. The items were – Tetracycline eye ointment, Amoxicillin, Paracetamol, Cotrimoxazole, Iron and Folic acid, Vitamin A, and ORS. Table A2 shows the impact of RBF on the availability of select drugs. In elevated facilities, only the availability of oral contraceptive pills increased by 22 percentage points, whereas there was no significant relative change among others.

Table A1: Effect of RBF pilot on facility infrastructure

	<i>Mean at baseline</i>		<i>Mean at endline</i>		Impact estimate	p-value
	Enhanced	Standard	Enhanced	Standard		
Facility has electric power	0.83	0.78	1.00	1.00	-0.048	0.689
Facility experiences no power outage	0.47	0.43	0.87	0.89	-0.067	0.668
Facility experiences no water outage	0.82	0.72	0.90	0.85	-0.026	0.894
Facility has functioning two-way radio	0.52	0.30	0.65	0.42	0.017	0.905
Facility has phone line	0.17	0.37	0.39	0.41	0.180	0.164
Facility has patient transportation means	0.61	0.59	0.74	0.73	-0.009	0.963
Facility has general outpatient consultation room	0.91	0.85	1.00	0.96	-0.024	0.723
Facility experiences no stock-out of disinfectant	0.65	0.85	0.95	0.81	0.326*	0.058
Facility has functioning incinerator for medical waste	0.26	0.30	0.43	0.41	0.063	0.763
Infrastructure index	-0.25	-0.13	0.32	0.03	0.406	0.460

Sample size 100 rural health centres; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

Table A2: Effect of RBF pilot on availability of drugs

	<i>Mean at baseline</i>		<i>Mean at endline</i>		Impact estimate	p-value
	Enhanced	Standard	Enhanced	Standard		
Paracetamol tabs	0.71	0.90	0.78	0.89	0.123	0.417
Amoxicillin tabs	0.71	0.90	0.74	0.81	0.126	0.374
Iron tabs	0.81	0.90	0.96	0.93	0.143	0.311
Folic acid tabs	0.85	0.81	0.74	0.88	-0.191	0.294
Cotrimoxazole	0.86	0.84	0.74	0.69	-0.018	0.884
Vitamin A	0.80	0.70	0.61	0.65	-0.170	0.433
Oral contraceptive pills	0.90	0.95	0.83	0.65	0.216**	0.032
Implant	0.81	0.60	0.70	0.77	-0.267	0.342
Artemisinin-Based Combination Therapy (ACT)	0.76	0.83	0.91	0.96	-0.009	0.946
Rifampicin	0.75	0.71	0.83	0.78	0.043	0.877
Magnesium sulphate	0.44	0.40	0.74	0.54	0.225	0.339
Misoprostol	0.60	0.77	0.35	0.27	0.143	0.464
Oxytocin	0.89	0.67	0.83	0.92	-0.308	0.333
Pentavalent vaccines	0.75	0.81	0.77	0.86	-0.039	0.872
Malaria rapid diagnostic kits	0.72	0.89	0.78	0.85	0.114	0.613
HIV testing kit	0.89	0.82	0.87	0.84	-0.072	0.741
Pregnancy testing kit	0.59	0.44	0.30	0.24	-0.168	0.602
Urine Dipstick	0.72	0.59	0.83	0.56	0.125	0.602
Drug availability index	-0.01	0.13	0.18	0.00	0.354	0.464

Sample size 100 rural health centres; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

Availability of functional medical equipment was assessed through direct observation of outpatient, sterilization, vaccination, delivery and neonatal equipment. An equipment availability index was constructed by assigning equal weight to the individual items and were further standardized. Individual items were children's weighing scale, height measure, tape measure, adult weighing scale, blood pressure instrument, thermometer, stethoscope, fetoscope, otoscope, and ambubag. Table A3 below shows there was no significant difference for any equipment.

Table A3: Effect of RBF pilot on availability of medical equipment

	<i>Mean at baseline</i>		<i>Mean at endline</i>		Impact estimate	p-value
	Enhanced	Standard	Enhanced	Standard		
Children's weighing scale	0.96	0.93	0.09	0.04	0.019	0.850
Height measure	0.52	0.37	0.09	0.00	-0.064	0.576
Tape measure	0.43	0.63	0.04	0.07	0.164	0.323
Adult weighing scale	0.96	0.93	0.09	0.04	0.019	0.869
Blood pressure instrument	0.78	0.78	0.04	0.00	0.039	0.722
Thermometer	0.96	0.85	0.04	0.04	-0.098	0.477
Stethoscope	0.87	0.81	0.00	0.11	-0.166	0.189
Fetoscope	0.70	0.89	0.04	0.04	0.200	0.185
Otoscope	0.22	0.19	0.00	0.00	-0.032	0.777
Electric autoclave (pressure and wet heat)	0.00	0.15	0.04	0.04	0.155	0.108
Refrigerator	0.87	0.70	0.17	0.15	-0.140	0.406
Delivery table/bed	0.74	0.63	0.00	0.00	-0.110	0.464
Partograph	0.48	0.44	0.00	0.04	-0.071	0.534
Baby scale (infant weighing scale)	0.65	0.78	0.00	0.04	0.089	0.475
Forceps, artery	0.65	0.67	0.00	0.04	-0.023	0.823
Needle holder	0.61	0.52	0.00	0.00	-0.090	0.655
Bag Valve Mask (Ambu bag)	0.17	0.15	0.00	0.07	-0.100	0.352
Guedel airways-neonatal, child, and adult	0.04	0.07	0.00	0.00	0.031	0.690
Equipment availability index	-0.01	-0.03	0.25	0.33	-0.107	0.819

Sample size 100 rural health centres; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

RBF facilities are supervised every quarter by the district hospital, which monitors the quality of services using a quality checklist. This checklist has 10 dimensions, each consisting of several items, weighted accordingly to their perceived importance to service delivery. The quality checklist used by the district hospital and the facility survey instrument used during data collection for this impact evaluation contain a few common items. These common items were extracted from the facility survey and the weights from the quality checklist applied, to construct a quality index similar to that used by the district hospital. Standardized indices were constructed for each quality dimension. Only four out of ten quality dimensions on the quality checklist could be mapped to the health facility instrument. Table A4 summarizes the results of the quality mapping exercise. Enhanced facilities report a decrease in the supply management index by 1.1 SD at 10% level of significance.

Table A4: Effect of RBF pilot on structural quality (mapping of quality checklist to facility survey)

	<i>Mean at baseline</i>		<i>Mean at endline</i>		Impact estimate	p-value
	Enhanced	Standard	Enhanced	Standard		
Curative Care	-0.005	0.004	-0.073	0.059	-0.122	0.612
Delivery Room	0.174	-0.141	0.381	-0.308	0.374	0.454
Supply management	0.029	-0.024	-0.579	0.469	-1.101*	0.096

Sample size 100 rural health centres; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

A1.1.2 Process quality

This section summarizes the effect on quality of care provided for antenatal, postnatal and child health care services. There are two main sources for this data: direct clinical observations of consultations and exit interviews administered to patients as they are leaving their consultations.

Maternal care

Table A5 presents similar quality measures for ANC, measured using patient exit interviews conducted as patients left care facilities. Participants were asked a series of questions as to whether certain ANC services were performed during their visit. The items were summed to create a composite score for the number of ANC services that were performed. More clients reported to have had their blood samples collected (23 percentage points) and pregnancy danger signs explained (23 percentage points) in the elevated facilities as compared to the standard facilities.

Table A5: Effect of RBF pilot on quality of antenatal care: results from patient exit interviews

	<i>Mean at baseline</i>		<i>Mean at endline</i>		Impact estimate	p-value
	Enhanced	Standard	Enhanced	Standard		
Weighed	0.81	0.90	0.93	0.87	0.150	0.173
Blood pressure measured	0.57	0.72	0.87	0.85	0.165	0.189
Urine sample collected	0.17	0.18	0.23	0.17	0.078	0.527
Blood sample collected	0.57	0.60	0.69	0.50	0.216**	0.011
Abdomen measured	0.33	0.40	0.37	0.50	-0.052	0.697
Abdomen palpated	0.87	0.86	0.95	0.94	-0.009	0.830
Advice on diet	0.45	0.58	0.71	0.69	0.152	0.126
Given/prescribed iron folic acid	0.85	0.90	0.97	0.95	0.061	0.492
Side effects of iron folic acid explained	0.17	0.19	0.10	0.17	-0.064	0.570
Given/prescribed anti-malarials	0.75	0.74	0.81	0.89	-0.084	0.380
Explained danger signs of pregnancy	0.40	0.63	0.75	0.75	0.230**	0.029
Quality of ANC index	-0.19	0.10	-0.07	-0.01	0.228	0.386

Sample size 525; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

Child health care

Quality of child care was measured both through patient exit interviews and through direct clinical observations. Participants were asked a series of questions as to whether certain child care services were conducted during the visit. The six items were summed to create a composite score for the

number of child care services that were performed. The patient exit data show no significant differences in any of these variables (Table A6).

Table A6: Effect of RBF pilot on quality of child health care: results from exit interviews

	<i>Mean at baseline</i>		<i>Mean at endline</i>		Impact estimate	p-value
	Enhanced	Standard	Enhanced	Standard		
Asked age	0.84	0.91	0.91	0.93	0.046	0.403
Weighed child	0.56	0.64	0.38	0.44	0.019	0.846
Measured height	0.07	0.10	0.03	0.05	0.002	0.968
Plotted a growth chart	0.33	0.39	0.07	0.16	-0.042	0.747
Physically examined	0.71	0.63	0.55	0.66	-0.183	0.205
Quality of care index	0.73	0.80	0.60	0.77	-0.112	0.323

Sample size 256; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

A1.1.3 Client satisfaction

This section summarizes the effect of the intervention on client satisfaction in antenatal care and child health care. Participants were asked to rate if the overall quality of the services was satisfactory on a Likert scale (1=strongly disagree to 5=strongly agree). The results in Table A7 do not show any significant differences in satisfaction on antenatal care. However, the results in Table A8 on child health care show lower proportion of clients who reported to be satisfied on cleanliness (22% points) and staff attitude (18% points) in the enhanced facilities as compared to the standard facilities.

Table A7: Effect of RBF pilot on client satisfaction in antenatal care: results from exit interviews

	<i>Mean at baseline</i>		<i>Mean at endline</i>		Impact estimate	p-value
	Enhanced	Standard	Enhanced	Standard		
The health facility is clean	0.80	0.75	0.84	0.93	-0.126	0.109
The health staff are courteous and respectful	0.85	0.85	0.84	0.86	-0.039	0.689
The amount of time you spent waiting to be seen by a health provider was reasonable	0.69	0.73	0.66	0.80	-0.106	0.372
The health worker spent a sufficient amount of time with the patient	0.82	0.82	0.92	0.94	-0.019	0.810
The hours the facility is open is adequate to meet the needs of the community	0.82	0.75	0.90	0.93	-0.082	0.462
All in all, you trust the health worker completely in this health facility	0.84	0.85	0.92	0.93	0.000	0.997
Satisfaction index	0.01	-0.10	-0.34	-0.01	-0.435	0.226

Sample size 525; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

Table A8: Effect of RBF pilot on client satisfaction on child health care: results from exit interviews

	<i>Mean at baseline</i>		<i>Mean at endline</i>		Impact estimate	p-value
	Enhanced	Standard	Enhanced	Standard		
The health facility is clean	0.85	0.79	0.77	0.92	-0.217*	0.053
The health staff are courteous and respectful	0.78	0.76	0.74	0.88	-0.179**	0.023
The amount of time you spent waiting to be seen by a health provider was reasonable	0.64	0.78	0.63	0.70	0.037	0.803
The health worker spent a sufficient amount of time with the patient	0.77	0.83	0.83	0.91	-0.029	0.817
The hours the facility is open is adequate to meet the needs of the community	0.76	0.77	0.78	0.93	-0.149	0.218
All in all, you trust the health worker completely in this health facility	0.76	0.79	0.85	0.95	-0.090	0.350
Satisfaction index	-0.10	-0.07	-0.47	0.03	-0.538	0.142

Sample size 256; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

A1.2 Effect on the health system

A1.2.1 Facility governance and autonomy

As shown in Table A9, the health centre committees seemed to be more active in enhanced facilities as they reported significantly higher number of meetings (1.2 more meetings per year on average) and external staff performance assessment linked to staff salaries or bonuses. However, enhanced facilities reported a lower probability of supervision visits by DHMT.

Facilities were interviewed about the level of autonomy during the follow up survey. The questions were related to the perceived autonomy of the facility in-charge on assigning task to staff, allocating budget, provision of services, and obtaining resources. The responses were recorded on a Likert Scale with values ranging from 1 (least autonomy) to 5 (maximum autonomy). An autonomy index was constructed utilizing select elements of autonomy such as ability to allocate resources and tasks effectively within the facility. The index was further standardized. As shown in Table A10, there were no significant differences in any of these variables.

Table A9: Effect of pilot RBF on facility governance

	<i>Mean at baseline</i>		<i>Mean at endline</i>		<i>Impact estimate</i>	<i>p-value</i>
	<i>Enhanced</i>	<i>Standard</i>	<i>Enhanced</i>	<i>Standard</i>		
Facility has a Hospital/Health Center Committee	0.87	0.93	1.00	1.00	0.056	0.626
Number of members on this Committee	13.89	12.25	11.36	12.54	-2.969	0.166
Number of Health Center Committee meetings held in the last 12 months	3.74	3.96	5.17	4.36	1.203**	0.039
Facility has written records of the Hospital/Health Center Committee meetings	0.65	0.80	0.83	0.78	0.188	0.270
Facility has a workplan for the current financial year	0.50	0.64	0.74	0.67	0.218	0.116
Number of visits made by a district hospital representative for supervision	0.61	1.52	1.39	1.26	1.042	0.105
Number of visits made by the District Health Management Team for supervision	1.52	1.63	1.74	2.78	-0.931*	0.093
Number of visits made by the local government for supervision or technical support	0.22	0.11	0.22	0.81	-0.704	0.114
Number of times performance of staff assessed internally	2.30	2.42	5.52	4.85	0.678	0.737
Internal staff performance assessment linked to staff salaries or bonuses	0.00	0.16	0.58	0.65	0.153	0.401
Number of times performance of staff assessed externally	1.74	2.12	2.30	4.22	-1.581	0.102
External staff performance assessment linked to staff salaries or bonuses	0.00	0.17	0.59	0.36	0.426** *	0.009
Number of times performance of the facility as a whole assessed externally	1.50	2.08	5.17	3.15	2.664	0.106

Sample size 100 rural health centres; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

Table A10: Effect of RBF pilot on facility autonomy

	<i>Mean at baseline</i>		<i>Mean at endline</i>		Impact estimate	p-value
	Enhanced	Standard	Remote	Standard		
Able to allocate my facility budget	0.88	0.74	0.94	0.89	0.158	0.536
Able to assign tasks and activities to staff	1.00	0.89	0.94	0.89	-0.000	0.226
DHMT supports my decisions and actions for doing a better job in my facility.	0.88	0.84	0.88	0.95	0.105	0.495
Choice over who I allocate for what tasks.	0.75	0.74	0.75	0.84	0.105	0.471
Choice over what services are provided in the facility.	0.69	0.58	0.44	0.47	-0.105	0.576
Enough authority to obtain the resources I need (drugs, supplies, funding)	0.81	0.68	0.75	0.63	-0.053	0.966
policies and procedures for doing things are clear to me.	0.88	0.95	0.88	0.95	0.000	1.000
DHMT provides adequate feedback to me about my job and the performance of my facility.	0.63	0.79	0.94	0.89	0.105	0.263
Autonomy index	0.04	-0.05	0.23	0.25	0.300	0.782

Sample size 100 rural health centres; * p<0.1 ** p<0.05 *** p<0.01; impact estimates with standard errors clustered at district level

A1.2.2 Satisfaction and motivation of the health workers

This section outlines the change in job satisfaction and motivation of health workers in elevated RBF health facilities (where prices of RBF indicators were 25% more than the normal prices of RBF indicators) compared with standard RBF health facilities (those that were earning normal prices on RBF indicators). As shown in table A11 and A12, there is no significant change in health worker motivation and satisfaction.

Table A11: Effect of RBF on job satisfaction

	Enhanced vs. standard	
	N	β (s.e.)
Relationship within facility	130	-4.354 (7.885)
Relationship outside of facility	130	2.519 (5.130)
Work conditions	130	2.813 (5.709)
Compensation	130	-3.909 (9.075)
Recognition	130	3.095 (5.827)
Opportunities	130	0.512 (4.890)

Facility fixed effects adjusted for age, sex, and cadre; SEs clustered at district level; * p<0.1 ** p<0.05 *** p<0.01

Table A12: Effect of RBF on work motivation

	Enhanced vs. standard	
	N	β (s.e.)
Teamwork	130	0.775 (4.666)
Autonomy	130	-0.681 (7.131)
Changes in facilities	130	-1.922 (2.094)
Work environment	130	-7.190 (4.863)
Self-concepts	130	-4.552 (3.748)
Recognition	130	1.593 (5.492)
Well-being	130	-1.755 (2.896)
Leadership of facility	130	-6.632 (4.972)

Facility fixed effects adjusted for age, sex, and cadre; SEs clustered at district level; * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

A1.3 Conclusions and discussions

While RBF mechanisms have demonstrated effectiveness in various settings, including to some degree in Zambia, there is an open question as to which components of RBF are most effective. One way to test whether it is the resource channel as powered through higher incentive levels that can be leveraged for even greater gains, the Zambian RBF evaluation included a direct test of elevated incentive levels on facility outcomes by randomizing a subset of remote facilities to receive 25% higher incentive payments.

We ran several statistical tests for service provision infrastructure, quality measures, health system, and functionality of health workers. On structural quality, we found no significant differences between enhanced and standard facilities for several indicators of structural quality except for increased availability of oral contraceptive pills in the elevated health facilities. However, there was a decrease in the availability of disinfectants and the overall supply management index in elevated facilities as compared to standard health facilities. Elevated health facilities performed better in process quality for antenatal care services. More clients reported to have had their blood samples collected and pregnancy danger signs explained in the elevated facilities as compared to the standard facilities. Nonetheless, there were no significant differences in the variables for process quality for child health care services.

Regarding governance and managerial autonomy, health centre committees in elevated health facilities were more active, while external staff performance assessments linked to staff salaries or bonuses were conducted regularly. However, elevated facilities reported a lower probability of supervision visits from their DMOs. Infrequent supervision visits to remote health facilities could be as a result of the long distances and bad road terrain between the DMOs and remote health facilities. However, one would have expected the DMOs to have been visiting all the health facilities in their district regardless of the facility being remote or not because the DMOs were being paid a performance payment linked to supervision.

The study examined perceived autonomy of the facility in-charge on assigning task to staff, allocating budget, provision of services, and obtaining resources. The results showed no significant differences

in any of these variables between the elevated and standard health facilities. This could be due to the fact that all of the variables which were examined were not affected by differentiated incentive payments as both elevated and standard health facilities were receiving RBF payment which were tied to outputs. The data also showed no significant differences in health worker motivation and satisfaction between elevated and standard health facilities.

Based on the results, we conclude that higher incentive payments in remote areas has a positive effect on a few measures from many considered indicators for process quality for antenatal care services, functionality of health centre committees, and external staff performance assessments. We found no significant differences between elevated and standard facilities for several indicators on managerial autonomy, and motivation and satisfaction of health workers. On the other hand, standard health facilities performed better on overall supply management, client satisfaction for child health services, and a higher probability of supervision visits from their DMOs.

Taken together we conclude that differences in the prices of the RBF indicators by 25% between the elevated and standard health facilities might not have yielded significant enough RBF payments and personal monetary incentives large enough to influence a lot of change. An alternative interpretation consistent with the findings is that gains under RBF programs are not particularly sensitive to incentive levels when they are not applied differently, and other aspects of the programme – the cognitive salience of the incentives (independent of amount), and monitoring and supervisory feedback, are also effective channels for achieving health gains.

In other RBF programs in Africa, the remoteness bonus was pegged at 30%²⁴ of the total quantity amount rather than having a higher price for each indicator. We also observe that remote health facilities had fewer members of staff and low catchment populations as compared to standard health facilities. Thus, despite the prices of all the quantity indicators in the elevated health facilities being 25% more than the normal prices for the indicators, the total amount of RBF incentive payments in remote health facilities were still much lower as compared to the standard health facilities. Furthermore, while we didn't examine the income effect on overall population health and health seeking behaviour, this could be another factor which can determine how much a health facility earns.

One main recommendation is to make sure that future RBF remoteness bonus allocation criterion go beyond the use of distance between the DMOs and health facility, and state of roads to include other factors related to care-seeking: the average distance between the health facility and communities, access to communication, availability of public transportation, catchment population, area of coverage, and population density.

²⁴ Ministry of Health and Child Care (2014). Zimbabwe National Results Based Financing Approach: Programme Implementation Manual

Appendix 2

Research Question 3 – How does the likelihood of audit/external verification of results affect the accuracy of reported data?

To answer the above question, a counter external verification exercise was conducted. This evaluation looked at the accuracy of data, and state of record maintenance at health facilities in the RBF arm. Health facilities in the RBF arm were randomized into three groups and each group was given a letter notifying them of a 100%, 30%, or 10% likelihood of data audit at the start of the RBF project. The hypothesis was that health facility managers who are absolutely certain that auditors or external verifiers will visit their health facilities were more likely to maintain good quality data and accurate records.

B1.1 Data requirements, sources, and selection of health facilities

Data was collected for the entire year (2013) on all the nine (9) incentivized RBF indicators. The key sources of data at the health facilities were the Health Information Aggregation (HIA) 2 forms in which health facilities summarized services provided for each indicator. Since this data required verification, tally sheets, activity sheets, and registers were used as these indicated the individual services delivered to a single client. This was also relevant in that it provided further details as to the date of the service, client register number, name of client, residential address, and other information. Hence, this data was used to check errors relating to summing, recording and data entry.

Since the overall objective of the counter-external verification exercise was to determine whether the likelihood of audit or external verification affected the accuracy in reporting, health facilities were grouped into three (3) categories as shown in table B1 below. Based on the list of health facilities with assigned probabilities or likelihood of audit, the counter-external verification team proceeded to apply simple random sampling to each category of health facilities in Excel® and obtained thirty-five (35) from each subgroup. Hence, data was collected from 105 health facilities (out of 176 health facilities on the RBF IE) from all the nine (9) provinces on the RBF project (Southern, Central, Copperbelt, Eastern Western, North Western, Luapula, Northern and Muchinga).

Table B1: Sampling in the Quality of Care study

	Likelihood of Audit		
	100%	30%	10%
Health facilities across the 10 RBF districts (population)	38	86	52
Number of Health Facilities sampled	35	35	35

Additional data was collected by the formal external verification exercise conducted by the international NGO Eurohealth and contracted by the Government of Zambia. This audit activity collected information from 72 health centres according to the announced audit likelihoods. Similar

to the counter-external verification, the official verification team conducted reconciliation between facility tally sheets and the HIA forms.

B1.2 Results

The first lesson from this experiment is that, when introducing the concept of data verification in a health system with little previous experience of these activities, repeated outreach to facility management combined with experiential learning may be necessary to internalize the meaning of a verification audit. This also applies to the possible sanctions for mis-reporting deemed deliberate.

What is clear from the formal external verification is that facility in-charges, for the most part, failed to internalize the announced likelihood audits – these were announced at RBF training and then in two follow-up letters to each in-charge. Table B2 shows the rate of understanding of audit likelihood among in-charges at time of survey. It is strikingly low – both for awareness of audit likelihood (as conveyed in official letters) and for discussion of audit likelihood with relevant facility staff.

Table B2: Facility in-charge awareness of audit likelihood

	% Yes
Did facility receive original announcement of audit likelihood?	23.6%
Was the original announcement discussed with appropriate staff?	22.2%
Did facility receive reminder announcement?	34.7%
Was reminder announcement discussed?	31.9%

Based on responses from 72 facilities, facility in-charge respondent for 53 facilities, proxy respondent for remaining 19

Most likely as a result the audit likelihoods are not related to accuracy of reported figures. No measure of divergence between the services reported in the tally and activity sheets and the services reported in the HIA 2 form is related to the announced audit likelihood. The third research question of the IE was not able to be answered as posed, as the announced programme does not appear to have been sufficiently comprehended by the study subjects.

However, a parallel question relating to the accuracy of reported information under RBF can be investigated as the counter external verification also collected reporting information from a sample of C2 facilities, where self-reported indicators of service provision were not incentivized as in RBF districts.

In the counter external verification, the quantitative analysis of internal consistency of facility data was done facility-by-facility and indicator-by-indicator, on all monthly observations. This data is presented in Table B3 which shows the accuracy of reporting by mean verification factor per indicator and the associated p-value. The mean verification factor (VF) was arrived at by using the following formula:

$$VF = \frac{\text{figures verified from tally and activity sheets (other registers)}}{\text{figures in HIA 2 forms (reported figures)}}$$

A verification factor greater than one (1) implies under-reporting while a verification factor lower than one (1) implies over-reporting. The first outcome investigated is whether the facility under-reported or over-reported the indicator over the 2013 calendar year as a function of residing in an RBF district. In other words, the first outcome is whether the VF is greater than one or less than or equal to one, and this outcome is regressed on a binary indicator for whether the reporting facility is in an RBF district.

Table B3: Facility reconciliation between reported services and tally sheet totals for nine related RBF indicators

Indicator	Percentage of control facilities under-reporting	Relative likelihood of RBF facilities underreporting	P-value of relative likelihood
Out-patient visit	0.643	0.049	0.634
Delivery	0.833	-0.154	0.100
Ante-natal care	0.654	0.029	0.787
Post-natal care	0.565	0.030	0.796
Immunization	0.588	-0.033	0.804
IPT	0.615	0.016	0.881
Family planning methods	0.552	-0.018	0.864
HIV testing and counselling	0.655	-0.093	0.369
PMTCT	0.875	-0.156	0.082

Based on counter-verification audits for all months in 2013, 140 total facilities - 35 in C2 districts and 105 in RBF districts

The results show that there was under-reporting in all the nine (9) indicators in the control facilities, ranging from 83% of facilities under-reporting the number of deliveries to 55% under-reporting for modern family planning usage. In terms of RBF facilities, the rate of under-reporting was indistinguishable from control facilities for 7 of the 9 indicators assessed. For two indicators – deliveries and PMTCT – the rate of underreporting was attenuated at moderate levels of significance (p-values of .10 and .08). Both of these indicators were 15 percentage points less likely to be under-reported. In sum, the reporting errors (which were negative errors on average) were largely the same for RBF and non-RBF facilities indicating that the tendency to inflate service counts of incentivized services are largely absent in the Zambia RBF. For the 2 indicators that are marginally less likely to be underreported in RBF districts – institutional deliveries and PMTCT – the mean reporting rates do not exceed the tally sheet total but simply come closer to this figure.

Besides the binary measure for under-reporting, the magnitude of reporting error can also be investigated with the data. This entails a regression of the magnitude of the VF on a binary indicator of RBF status (results not shown). Since the VF is a ratio, the magnitude of under-reporting and over-reporting errors are not symmetric. For this reason, separate regressions are run for VFs > 1 and VFs ≤ 1. In terms of the magnitude of under-reporting (VFs >1), there is no identified systematic relationship with RBF status. For those facilities that over-report (VFs ≤1), the magnitude of over-report is actually significantly attenuated in RBF facilities for four indicators at traditional levels of significance: outpatient consultations, immunizations, IPT, and HIV testing and counselling of pregnant women (p-values of .02, .01, .04, and .03 respectively). It appears that when reporting behaviours of RBF facilities differ from control facilities, it is to bring reported services closer to the quantity of services recorded in the activity sheets.

B1.3 Conclusions and discussion

We hypothesized that the likelihood of audit or external verification of results affects the accuracy of reported data. The thinking was that health facility managers who are absolutely certain that auditors or external verifiers will visit their health facilities were more likely to have accurate data, and high quality records. To test our hypothesis, health facilities in the RBF arm were randomized into three groups and each group was given a letter notifying them of a 100%, 30%, or 10% likelihood of data audit. These letters were issued at the start of the RBF project and a counter external verification exercise was conducted at the end of the RBF project. Unfortunately, it doesn't seem that this audit experiment resulted in a usable analysis as most facility in-charges were unaware of the relative audit likelihoods despite repeated outreach efforts.

However, the data can be used to address a general issue of reporting veracity under an RBF programme. One hypothesis is that the introduction of incentives may bring about gaming behaviour that inflates the self-reported service totals over that which was actually supplied by health clinics. Instead the data shows that under-reporting of services is widespread in control facilities and that, if anything, RBF facilities report more accurately, at least with respect to select indicators, without appreciable over-reporting.

In view of the above, we found no evidence that incentive introduction under RBF affects the accuracy of reported data except in so far as to make it more accurate. The precise explanation for this was not established but one possible cause may be that most facilities did not have dedicated staff for managing the data and so, not all records were utilized in aggregating the primary entries for reporting. Therefore, record keeping and storage of files was at best a haphazard affair in which some registers were not considered, for example, the large number of beneficiaries of certain services such as curative consultation required more than one register being used within a month. Although the condition of records varied between facilities, it was observed that those facilities that had designated data entry clerks had better organised records than those without. Facilities with poor or incomplete records attributed their inadequacy to work overload as their officers did not have adequate time outside their normal duties to complete records. To the extent that RBF prioritizes the use of data in facility compensation, it actually promotes more accurate self-reported information systems.

Appendix 3

Table C1. Mapping of facility questionnaire instrument into the balanced score card

CHECK LIST ITEMS	CRITERIA	POSSIBLE MAXIMUM SCORE	Question # in IE facility survey checklist
CURATIVE CARE			
Equipment available in the treatment room and in working condition: 1) Thermometer 2) Blood pressure machine 3) Stethoscope 4) emergency trolley set 5) Unused non-sterile gloves 6) Adult scale 7) Sharp boxes 8) Examination table 9) Running water from tap or bucket with Tap filled with water for washing hands with soap	Treatment room equipped with 9 functional materials = 9 One material missing or non-functional = deduct 1 per missing element More than five material missing or non-functional = 0	9	Q1.45 Q1.47 Q13.02
Documentation for consultation available to provider: 1) Integrated Treatment guidelines (ITG), 2) Tally sheets 3) IMCI guidelines available	2 documents present in consultation room = 3 1 documents missing = deduct 1	3	Q12.19
FAMILY PLANNING			
(If not on clinic day), Analysis of 10 randomly chosen FP clients from the past three months: number of entries completely filled in all fields	Yes = 1.5 No = 0	15	Q7.17
DELIVERY ROOM			
Available and functional equipment and supplies: 1) Adjustable, clean delivery table 2) at least 3 sterilized instrument boxes (with needle holder, two Kocher clamps, toothed forceps, two pairs of scissors) 3) Neonatal suction devise 4) Foetal scope 5) Suture thread, 6) light source 7) infant weighing scale 8) Sterilizing drum 9) Ophthalmic ointment 10) Gauze drum 11) Plastic apron 12) local anesthesia (at least 20ml in reserve) 13) unused and non-torn surgical gloves 14 umbilical cord clamp	One material available and functional = 1 If more than 50% of elements not available = 0	14	Q13.09
SUPPLY MANAGEMENT			
Cleanliness of pharmacy (no dust on shelves and products, no cobwebs)	Cleanliness assured = 2 Cleanliness not assured= 0	2	Q14.06
Stocking in accordance with regulations: 1) Products arranged on shelves, not on floor 2) Logically arranged products (alphabetical order or by type of therapy) 3) On basis of expiry date 4) With signs on shelves according to International Common Denomination (Generic names) 5) Agreement between theoretical and physical stock	Stocking fulfilling all criteria = 5 each unmet criterion = 0	5	Q14.09

Appendix 4

Population outcomes and Fisher-exact standard errors

The evaluation design for the Zambia RBF pilot involved pair-matched randomization at the district level. Randomization at the district level does come with potential inferential cost in the power of the analysis as the number of units of randomization is limited. In the case of the Zambia RBF, the RBF and C1 interventions were each piloted in 10 districts around the country. However due to budgetary limitations, population data was only able to be collected in six districts in each study arm. The main report estimates standard errors for impact estimates by clustering at the PSU level. However, as implied above, there may be unobserved influences at the district level that lead to district-level correlations in impacts that ideally should be accounted for. However, this presents two analytic difficulties.

The first difficulty is relatively few study units for the analysis. Besides the challenge to inferential power by the relatively few number of study units, traditional approaches to standard error estimates, notably the cluster-robust standard error, may be downward biased and thus over-reject the null hypothesis of no treatment effect (Cameron et al., 2008). To counteract this potential bias, the precision of statistical tests can also be assessed through Randomization Inference (RI) which assumes all observed outcomes and covariates to be fixed and generates the reference distribution of test statistics by modelling the treatment assignment as the sole random variable in the data (Ernst, 2005). RI compares the actual test statistic observed in the evaluation against the distribution of all conceivable test statistics as determined through permutation methods – where the actual statistic falls in this distribution determines the exact p-value. This one-tailed hypothesis test is considered an exact test because it does not require a large-sample approximation as randomization itself is the basis for inference and permutation methods have exhausted all possible treatment assignments across districts. An exact test has the added benefit that it does not impose distributional assumptions that are often behind approximations of reference distributions in standard hypothesis testing.

This appendix presents the population level impacts with exact p-values estimated through randomization inference and compares them with the asymptotic p-values estimated with clustering at the PSU level and that were reported earlier. The tables show that indeed the precision of the inference is not as great with the exact p-values. Many impacts that were estimated to be precisely estimated at traditional levels of statistical significance (with standard errors clustered at the survey cluster level) are no longer such. This raises the question of the acceptable level of precision for impacts to inform policy when an evaluation does not have high power. Given the that the first pilot of RBF in Zambia faced various implementation challenges, population data was unable to be collected on a broad basis, and the international evidence base for RBF mechanisms comprises only a handful of countries, policy makers may wish to consider exact p-values larger than traditional cut-off levels, say on the order of 0.15, as sufficiently precise to inform future policy directions.

Table D1: In-Facility delivery indicators

All births within two years of survey	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Facility delivery	-0.049	0.308	0.193	0.128***	0.115	0.005	0.175***	0.038	0.000
Skilled provider and facility	-0.043	0.374	0.304	0.101**	0.226	0.046	0.142***	0.144	0.007

Table D2: Antenatal care coverage

All pregnancies within two years	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Any ANC	-0.015	0.266	0.119	-0.015	0.263	0.210	0.000	0.493	0.970
ANC: proper provider and facility	0.052*	0.280	0.061	0.041	0.326	0.296	-0.014	0.442	0.730
4 or more ANC visits	-0.004	0.471	0.898	-0.034	0.293	0.320	-0.029	0.319	0.408
Number of ANC visits	-0.074	0.305	0.483	-0.187	0.093	0.118	-0.112	0.218	0.347
Timing of first ANC visit	-0.372***	0.096	0.000	-0.476***	0.043	0.000	-0.108	0.358	0.320

Table D3: Postnatal care coverage

All births within two years of survey	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Any PNC	-0.051	0.340	0.202	0.082**	0.249	0.042	0.132***	0.129	0.002
PNC: proper provider and facility	-0.010	0.473	0.798	0.103*	0.238	0.054	0.112**	0.219	0.036

Table D4: Family planning indicators

Women 15-49 in recent birth hhs	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Any Contraception	-0.024	0.380	0.582	-0.039	0.307	0.378	-0.017	0.416	0.714
Modern Contraception (of FP users)	-0.045	0.204	0.151	0.002	0.486	0.942	0.046	0.195	0.149
Any FP outreach	-0.078**	0.127	0.021	0.083***	0.110	0.005	0.159***	0.005	0.000

Table D5: Immunization Coverage for children aged 12-23 months

Children between 12 and 23 months	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Full vaccination	0.116	0.271	0.215	0.052	0.391	0.569	-0.046	0.405	0.587
Any vaccinations	-0.066	0.171	0.116	0.015	0.418	0.716	0.081**	0.115	0.043
BCG inject ever received	0.031	0.232	0.212	0.07**	0.027	0.020	0.028	0.252	0.248
Polio ever received	0.010	0.423	0.768	0.027	0.273	0.479	0.014	0.379	0.610
Polio received right after birth	0.003	0.484	0.970	0.067	0.317	0.422	0.062	0.331	0.449
DPT ever received	-0.010	0.418	0.718	0.061**	0.082	0.073	0.056	0.104	0.108
DPT freq given	0.098	0.354	0.441	0.168	0.259	0.198	0.083	0.374	0.553
Measles ever received	-0.053	0.231	0.304	-0.033	0.320	0.478	0.013	0.430	0.783

Table D6: Health seeking behaviour for general illness, separately for under-5s and over-5s

All children under 5	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Any sickness in previous 4 weeks	0.013	0.425	0.579	-0.007	0.461	0.798	-0.023	0.372	0.324
Sought advice for illness	0.052	0.188	0.456	-0.018	0.383	0.729	-0.065	0.134	0.300
Visited gov facility for illness	-0.029	0.286	0.550	-0.095**	0.027	0.042	-0.035	0.252	0.533
Visited any facility for illness	-0.030	0.238	0.442	-0.098**	0.002	0.010	-0.048	0.126	0.304
Visited CHW for illness	-0.009	0.426	0.804	0.054**	0.022	0.033	0.057	0.013	0.142
HH members over 5	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Any sickness in previous 4 weeks	-0.010	0.340	0.339	-0.022*	0.175	0.051	-0.012	0.307	0.297
Sought advice for illness	-0.038	0.346	0.630	-0.046	0.315	0.497	-0.009	0.464	0.900
Visited gov facility for illness	0.09*	0.084	0.072	0.037	0.303	0.529	-0.025	0.366	0.682
Visited any facility for illness	0.082*	0.095	0.067	0.039	0.281	0.476	-0.021	0.379	0.699
Visited CHW for illness	-0.054**	0.001	0.013	0.006	0.463	0.750	0.043	0.047	0.090

Table D7: Knowledge of maternal health danger signs: Results from the household survey

Women 15-49 in recent birth hhs	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Excessive Vaginal Bleeding	0.08**	0.237	0.037	0.045	0.343	0.275	-0.037	0.373	0.383
Fever	0.108***	0.186	0.005	-0.015	0.450	0.672	-0.12***	0.159	0.003
Swollen face, hands or legs	0.023	0.372	0.545	-0.055*	0.218	0.082	-0.082**	0.121	0.028
Difficulty breathing	0.030	0.270	0.167	0.005	0.456	0.815	-0.022	0.328	0.380
Sever headache	0.018	0.383	0.511	-0.041	0.245	0.149	-0.058*	0.169	0.082
Convulsions / Fits	-0.002	0.478	0.878	-0.035**	0.200	0.010	-0.033**	0.209	0.039
Lightheadedness / Dizziness	0.056**	0.158	0.042	0.014	0.412	0.650	-0.049*	0.194	0.097
Blurred vision	0.019	0.241	0.195	0.013	0.313	0.391	-0.005	0.427	0.748
High blood pressure	0.071**	0.165	0.023	0.07**	0.166	0.016	0.004	0.482	0.902
Sever pain in lower belly / tummy	0.164***	0.039	0.000	0.162***	0.041	0.000	-0.001	0.498	0.986
Water leaks or breaks	0.134***	0.043	0.000	0.12***	0.065	0.000	-0.018	0.415	0.514
Baby stops or reduces moving	0.078**	0.129	0.027	0.071**	0.154	0.042	-0.009	0.453	0.817

Table D8: Process quality of antenatal care provided: Results from the household survey

All pregnancies within 2 years of int.	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Blood pressure measured	-0.029	0.284	0.351	-0.028	0.294	0.327	0.000	0.498	0.995
Urine sample taken	-0.129***	0.064	0.000	-0.095**	0.138	0.017	0.031	0.363	0.425
Blood sample taken	-0.099***	0.161	0.008	0.057*	0.325	0.091	0.153***	0.031	0.000
Any tetanus injection	0.06**	0.123	0.031	-0.039	0.233	0.160	-0.1***	0.018	0.000
Number of tetanus injections	-0.041	0.384	0.595	0.067	0.307	0.400	0.109	0.201	0.143
Any iron taken	-0.001	0.494	0.973	0.044**	0.123	0.028	0.045**	0.119	0.013
Number of days iron taken	-16.56***	0.144	0.000	-15.30***	0.165	0.003	0.453	0.488	0.925
Malaria drugs taken	0.026	0.234	0.149	0.036**	0.149	0.019	0.010	0.392	0.567

Table D9: Process quality of postnatal care provided: Results from the household survey

All births within 2 years of survey	RBF vs. Control 1			RBF vs. Control 2			Control 1 vs. Control 2		
	Impact	Exact	p-value	Impact	Exact	p-value	Impact	Exact	p-value
Vit A given after delivery	-0.091**	0.177	0.023	0.025	0.404	0.485	0.115***	0.113	0.006
Ever breastfed infant	0.000	0.499	0.996	0.006	0.364	0.607	0.006	0.357	0.547
Immediately breastfed infant	-0.07*	0.288	0.053	0.136***	0.136	0.000	0.206***	0.041	0.000