An Impact Evaluation of the Results-based Financing Intervention in Afghanistan

FINAL REPORT 2015

KIT | The Royal Tropical Institute

SILK ROUTE
Training and Research Organization
موسسه آموزشی و پژوهشی راه ابریشم

CONTENTS

CONTENTS ii
LIST OF TABLES AND FIGURES iv
EXECUTIVE SUMMARY vi
ABBREVIATIONS x
BACKGROUND 11
Results-based financing pilot intervention in Afghanistan 11
Organization of this report 13
Ethical Approval 13
CHAPTER 1: HOUSEHOLD SURVEY RESULTS 14
Methodology 14
Survey design 14
Sample design 14
Questionnaires 15
Data collection fieldwork 16
Data management and analysis 17
Results 22
Household and resident characteristics 22
RBF Indicators on maternal and child health and equity 22
Descriptive Tables 26
CHAPTER 2: COMMUNITY LEADERS' SURVEY 35
Methodology 35
Sample 35
Data collection and management 35
Analysis 35
Results 37
CHAPTER 3: HEALTH FACILITY ASSESSMENT 43
Methodology 43
Sample 43
Data Collection 44
Comparison with 2009/10 BSC 44
Analysis 44

Results....46
9 province analysis....46
11 province analysis....48
CHAPTER 4: QUALITATIVE ASSESSMENT....51
SECTION 4.1: HEALTH FACILITY LEVEL....51
Methodology....51
Results....51
Community interaction....51
Health system referrals and increased work load....52
Health worker motivation....53
General RBF knowledge and perceptions....54
Targets, competition, and expectations among comparison group health facilities....55
Quality and quantity of services....55
RBF accountability mechanisms....57
Fund decisions and use....57
RBF verification....58
SECTION 4.2: COMMUNITY LEVEL....59
Methods....59
Results....60
Positive and negative changes in quality of health services....60
Challenges and opportunities....61
CHAPTER 5: DISCUSSION....62
CHAPTER 6: RECOMMENDATIONS....67
REFERENCES....69
ANNEX 1: Ex-post power calculation....71
ANNEX 2: Results using the original approach....74
ANNEX 3: Characteristics of dropped clusters....78
ANNEX 4: Baseline concentration curves....79

LIST OF TABLES AND FIGURES

Table 1 Results of the household and individual interviews at baseline 18

Table 2 Results of the household and individual interviews at endline 18

Table 3 Indicator definitions 21

Table 4 Selected household and resident characteristics* 22

Table 5 RBF indicators health indicators 23

Table 6 RBF equity indicators 24

Table 7 RBF indicators, before and after test* 25

Table 8 Current use of modern contraceptives: Percent distribution of currently married women who use modern contraception methods (N=6,293), weighted % 26

Table 9 Antenatal care coverage: Percent distribution of ever-married women who had at least one ANC visit to a skilled provider for the most recent live birth in the two years preceding the survey (excluding currently pregnant women) (N=2,085), weighted % 27

Table 10 Percent distribution of skilled birth attendance: Use of skilled attendants by respondent characteristics (N=2,268), weighted % 28

Table 11Percent distribution of postnatal care received by a skilled provider within 42 days of delivery by respondent characteristics (N=2,268), weighted % 29

Table 12 Institutional deliveries: Percent distribution of ever-married women who had a live birth in the two years preceding the survey with an institutional delivery for their most recent birth (N=2,268), weighted % 30

Table 13 Percent distribution of children 12-23 months who received PENTA3 vaccinations at any time before the survey by health card or mother's report, (N=1,038), weighted %. 31

Table 14 Percent distribution of children under five years of age who were sick two weeks before the survey, (N=5,935), weighted % 32

Table 15 Percent distribution of children under five years of age who were sick two weeks before the survey and sought outpatient care services, (N=2,633), weighted %. 33

Table 16 Percent distribution of children over five years of age and adults who were sick two weeks before the survey, (N=39,451), weighted %. 34

Table 17 Percent distribution of children over five years of age and adults who were sick two weeks before the survey and sought outpatient care services, (N=4,855), weighted % 34

Table 18 Sample of community leaders at baseline and endline 35

Table 19 Background characteristics of participants 37

Table 20 Direct observations (N= 262) 38

Table 21 Demographic characteristics (N = 262) 38

Table 22 Basic Services (N=262) 39

Table 23 Facilities or services available in village (N = 262) 39

Table 24 Women's rights (N = 262) 40

Table 25 Village governance and leadership (N = 262) 40

Table 26 Economic activity (N = 262) 41

Table 27 External shocks (N = 262) 42

Table 28 Description of sample for analysis 45

Table 29 Nine provinces – Matched-pair facility analysis from household survey sample 47

Table 30 Eleven provinces – Unmatched-pair facility analysis from total BSC sample 49

Table 31 Summary table of positive changes by study group 60

Table 32 Summary table of negative changes by study group 60

Table 33 Summary table of rankings of challenges 61

Figure 1 Concentration Curves for Equity Indicators of Utilization 25

EXECUTIVE SUMMARY

The Results Based Financing (RBF) pilot intervention was introduced in 2009 in 11 provinces in Afghanistan. The RBF scheme was integrated into the existing health system and implemented by existing health sector implementing partners targeting improvements in key maternal and child health indicators. Health facility workers were provided additional financial incentives for providing increased quantity and quality of services from baseline targets. Incentive payments to participating health facilities were expected to increase the quality of health services and quantity of services provided targeting women and children. The final impact evaluation consists of a household survey, community leaders' survey, health facility assessment, and a qualitative assessment of health facility staff.

Were there changes at the household level?

The household survey was conducted at baseline in 2010 in 9 of the 11 provinces and the follow-up household survey was conducted in 2015. The household survey aims to provide post-intervention information and baseline-endline comparisons on the coverage of health services on six health indicators, comparing households within the catchment areas of the intervention and comparison health facilities. At endline, 6,584 (3,300 in the treatment group and 3,284 in the comparison group) household interviews were completed with a nonresponse rate of less than five percent.

The key maternal and child health indicators assessed were current use of modern contraceptives; at least one antenatal care check up from a skilled provider; skilled birth attendant present at last delivery; postnatal care checkup within 42 days of delivery; children (12-23 months) received Pentavalent 3 vaccination; and equity in utilization of care for institutional delivery, child outpatient services, adult outpatient services, and adult morbidity.

Large improvements across all health indicators were made by both study groups since baseline. At endline, the treatment group has higher coverage of all indicators compared to the comparison group except in modern contraceptive use, but these differences are not statistically significant. In regards to equity of utilization of health services, the comparison group is more pro-poor at endline compared to the treatment group in their concentration indices, but testing of dominance of their concentration curves showed that there were no differences.

Were there changes at the community level?

A survey of community leaders was conducted at baseline to assess their perceptions of the community's demographic characteristics, economic activities, crises and other shocks, and social capital. At endline, leaders from 262 villages were interviewed, 130 from the intervention group and 132 from the comparison group. Overall, the intervention and comparison group groups were similar across village characteristics spanning economic activity, external shocks, women's rights, governance, and basic service availability. There were some indicators that showed statistically significant differences between the two study groups but most were unlikely to explain the RBF household survey

results. Poor infrastructure such as roads and lack of transportation were commonly mentioned as barriers to accessing care.

Were there changes at the health facility level?

The health facility assessment was designed to assess the impact of RBF on the performance of health facilities as measured by the Balanced Scorecard (BSC). Two analytical approaches were conducted. The first analysis presents results on the same nine province matched-pair health facilities randomly selected as part of the RBF household survey. The second analysis presents the results of all 11 RBF province health facility performance. The main domains assessed for health facility performance include Patients and Community, Staff, Physical Capacity, Service Provision, and Management Systems (for more detailed information on all performance indicators, please see Chapter 3).

The estimated difference in overall mean score between the intervention and comparison groups from baseline to endline was not statistically significant (at the 5% level) for the 11 province analysis and 1.8% difference for the nine province analysis ($p<0.01$). While the nine province analysis shows a difference that is statistically significant, it is not very large and possibly not truly meaningful in understanding the performance differences of the two study groups. When looking at the specific measures of health facility performance, the intervention health facilities have a statistically significant higher performance on seven indicators: engagement of community in decision-making, health worker satisfaction, staff received training, equipment functionality, health facility management functionality, pharmaceuticals and vaccines availability, and universal precautions.

What were the perceived changes, benefits, and consequences of RBF?

The qualitative assessment was conducted in three provinces: Badakhshan, Bamyan, and Parwan. A total of 39 in-depth interviews (IDI) and focus group discussion (FGD) were conducted with health facility staff, including heads of facility, administrative staff, and clinical health workers across all BPHS level health facility types. A total of 8 comparison group health facilities and 13 intervention health facilities were visited across the three provinces.

All health facility staff reported increased referrals of patients in terms of receipt of patients from lower level health facilities and referring out of patients to higher level health facilities. A strong network of CHWs and health posts in communities ensured referrals to the BPHS clinics – this was true for both study groups. All RBF intervention health facilities, mentioned the RBF payments as a key motivating factor, yet acknowledgment for performance and supportive supervision were important factors in motivating staff across study groups, regardless of financial payments. All comparison group health facility staff were aware of the RBF program and the program specific targets, but did not have a clear understanding of the randomized selection process. All treatment health facility staff showed higher knowledge of RBF details such as the randomized selection process, objectives, targets, and payment processes. Comparison group health facility staff expressed misunderstanding for why they were not included in the program and future expectations of participating in the program and receiving additional payments if they performed well and reached

their targets. Some health facilities expressed explicit expectations of receiving the program in the future, while others valued the importance of reaching their targets regardless of whether they had the RBF program.

Positive changes in the quality and quantity of health services over the past five years were reported by both RBF health facilities and comparison group health facilities. Both intervention and comparison group health facilities discussed how the routine HMIS verification changed health workers' behaviors to be more attentive to documentation and reporting of patients.

What are the limitations of this study?

While this study was designed as a randomized comparison group trial, which is considered a gold standard in assessing impact and attribution of an intervention, there are several limitations that should be considered. There is always potential for selection bias and response bias in a cluster sample. Additionally, due to insecurity, we were not able to visit 13 of the original clusters. However, the main limitation of this study is the lack of a true control group due to the design of the treatment assignment at the health facility type level rather than at the district level.

What can we conclude?

The results of the RBF household survey indicate that large improvements have been made in health indicators since baseline, but the intervention did not have a statistically significant positive effect overall. The RBF intervention seems to have had an overall positive effect at the health facility level (though slight) through the BSC as well as through health worker perceptions, yet these improvements have not translated to the population level. Possible reasons for the lack of treatment effect may be due to issues with fidelity of intervention, inherent issues in the study design, spillover and/or contamination and competition by the comparison group, unobserved changes in the comparison group, or the receipt of increased monitoring and supervision in the comparison group through HMIS verification. Recommendations include:

- Communicate clear targets to all health facilities implementing the BPHS and EPHS
- Maintain and improve the HMIS verification
- Improve monitoring and supportive supervision
- Consider alternative mechanisms for motivating health workers
- Ensure timely payment of salaries
- Consider alternative demand side interventions through the provision of transportation to promote use of care
- Conduct additional qualitative studies to improve the verification process

- Organize a stakeholder workshop to understand the RBF impact evaluation results in its entirety and to decide how to move forward

- Prepare a de-implementation plan and communication strategy if RBF is not to be continued or expanded

The MoPH is currently implementing the System Enhancement for Health Action in Transition (SEHAT) project, for which RBF was intended to be a promising financing mechanism for linking NGOs to health facility performance. This pilot project was shown to be ineffective for the various reasons discussed above. Health facilities seemed responsive to the HMIS verification process, which may be an alternative method to assess performance, rather than payments.

ABBREVIATIONS

ANC	Antenatal care
ANPHI	Afghanistan National Public Health Institute
BHC	Basic health center
BPHS	Basic Package of Health Services
CC	Concentration curve
CDC	Community Development Councils
CHC	Comprehensive health center
CI	Confidence interval
CPR	Contraceptive prevalence rate
DH	District hospital
DiD	Difference-in-difference
EPHS	Essential Package of Hospital Services
FGD	Focus group discussion
IDI	In-depth interviews
KIT	The Royal Tropical Institute
MDG	Millennium Development Goals
MoPH	Ministry of Public Health
P4P	Pay for performance
PBF	Performance-based financing
PENTA3	Pentavalent vaccine
PNC	Postnatal care
PPHD	Provincial Public Health Department
RBF	Results-based Financing
SC	Sub-health center
SE	Standard error
SEHAT	System Enhancement for Health Action in Transition
SRTRO	Silk Route Training and Research Organization

BACKGROUND

The Afghanistan health system has made significant improvements through the implementation of the Basic Package of Health Services (BPHS) and the Essential Package of Hospital Services (EPHS). The increased availability and access to primary and secondary health care services has resulted in a decrease in maternal and child mortality. Health outcome indicators have improved notably in recent years. By 2010, infant mortality decreased to 65 deaths per 1,000 births, under-5 mortality was down to 84 per 1,000 live births, and maternal mortality ratio was estimated at 327 per 100,000 live births from 129, 191, and 1600, respectively (1–3). Yet key maternal and child health services remained underutilized and the health status of the population remained among the poorest compared to other low-income countries.

Globally, the pay for performance (P4P), performance-based financing (PBF) or results-based financing (RBF) approach[1] was gaining recognition as a mechanism to increase the use, quality, and efficiency of health care services. RBF is a financing mechanism focused on outputs rather than inputs, by transferring money or goods conditional on predetermined performance targets (4). The financial incentive can operate on either the supply or demand side of health services and can be targeted to individual health care users, individual health care provider, health facilities, or public or private organizations delivering health care services. Experiences of pilot RBF programs as a supply-side financing mechanism have been shared from Rwanda (5), Nigeria (6), Burundi (7), Mali, and Ghana (8), among others. The design of RBF programs differ widely by country as it is context specific and used as a tool for reforming health systems. RBF schemes remain widely advocated for as a promising mechanism to achieve the Millennium Development Goals (MDGs), particularly in improving maternal and child health outcomes. Some countries have reported improvements in utilization of reproductive and maternal health services as well as increased quality of health care services (5,7). Studies have also reported undesirable effects of the incentive-based intervention such as gaming the system and widening inequities (9). Over 15 years of RBF programs, however, have resulted in limited evidence of its effectiveness due to the lack of rigorous evaluation studies in the published literature (10). This report provides the combined results of the household survey, health facility assessment, community leaders' survey, and qualitative assessment of the impact evaluation of the RBF pilot scheme in Afghanistan.

Results-based financing pilot intervention in Afghanistan

In 2009, the Ministry of Public Health with the support of the World Bank introduced a pilot results-based financing (RBF) scheme in 11 provinces[2] of the country. The RBF scheme was integrated into the existing health system targeting improvements in key maternal and child health indicators. The RBF intervention was implemented by existing BPHS implementing partners (NGOS and MoPH-SM). Health facility workers were provided additional financial incentives for providing increased

[1] Pay for performance, performance-based financing, and results-based financing are often used interchangeably. This report will use results-based financing to describe the supply-side intervention implemented in Afghanistan.
[2] RBF was implemented in the following 11 provinces: Badakhshan, Balkh, Bamyan, Jawzjan, Kandahar, Kunduz, Saripul, Samangan, Takhar, Parwan, Panjsher

quantity and quality of services from baseline targets. The quantity of services is reported through the HMIS, and the quality of services is measured through field monitoring. Verification of HMIS reports is conducted externally by a third party. Health facilities were matched by facility type by the number of outpatient visits in the year prior to implementation and randomly assigned to a treatment arm or comparison group arm. A total of 463 health facilities[3] in the 11 provinces are participating in the RBF pilot project, of which 246 health facilities are in the treatment arm and 217 health facilities are in the comparison group arm. Health facilities in the comparison group arm do not receive the RBF intervention but were followed throughout the same intervention period through HMIS data verification.

The objective of the RBF intervention is to help achieve MDG 4 (to reduce child mortality) and MDG 5 (to reduce maternal mortality) by implementing interventions that provide performance payments for health workers in order to:

1. Increase key maternal and child health outputs

2. Further improve the quality of health care services

3. Ensure that patients and communities are increasingly involved and satisfied with the publicly financed health services they receive

Goal and objectives of the RBF impact evaluation

The RBF impact evaluation consists of three parts: 1) household survey of 9 of the 11 RBF provinces;[4] 2) quality assessment of health facilities; and 3) qualitative study. The objectives of each component include the following:

- Household survey: To provide post-intervention information and baseline-endline comparisons on the coverage of health services at the community level to MOPH and its partners. This survey will produce estimates of the target 6 health indicators compared with baseline results.

- Community leaders' survey: To provide descriptive statistics on community-level development.

- Health facility assessment: To compare intervention and comparison group health facilities on quality score measures since baseline

- Qualitative assessment: In intervention health facilities, to understand challenges in the implementation of the RBF scheme, how the delivery of services has changed, and how health

[3] 115 Sub-health Centers (SC); 225 Basic Health Centers (BHC); 99 Comprehensive Health Centers (CHC); and 24 District Hospitals (DH)

[4] Two provinces (Kunduz and Kandahar) were dropped at baseline from the household survey due to insecurity, resulting in 9 provinces included in the household survey sample.

workers have been responsive to RBF. In comparison group health facilities, to understand general changes to health service delivery and quality and perceptions/knowledge of the RBF scheme.

Organization of this report

This report provides the methods and results of each component in separate chapters with an overall discussion and recommendations chapter. Chapter 1 presented the results of the household survey; Chapter 2 presents the results of the community leaders' survey; Chapter 3 presents the results of the health facility assessment; Chapter 4 presents the results of the qualitative assessment of health facility staff; Chapter 5 presents an overall discussion of these findings; and Chapter 6 presents recommendations.

Ethical Approval

We received ethical approval from institutional review boards (IRB) of the Afghanistan National Public Health Institute (ANPHI) and the Royal Tropical Institute (KIT) in April 2015. The impact evaluation of RBF was implemented by KIT and the Silk Route Research and Training Organization (SRTRO).

CHAPTER 1: HOUSEHOLD SURVEY RESULTS

This chapter presents the results of the RBF household survey. The baseline survey was conducted by Johns Hopkins University (JHU) in 2010. We surveyed the same clusters and households identified at baseline for the endline survey in 2015.

Methodology

Survey design

The 2015 endline household survey aims to report on the following indicators and their difference between the two study groups, and since the 2010 baseline household survey. The RBF indicators for assessment include:

- Proportion of currently married women of reproductive age who or whose partner are currently using at least one modern family planning method;
- Proportion of women who attended at least one ANC visit with a skilled provider for their most recent delivery;
- Proportion of births attended by a skilled attendant for a woman's most recent delivery;
- Proportion of children 12 to 23 months of age getting three doses of DPT/Pentavalent before their first birthday;
- Percentage of women who received at least one postnatal care from a trained provider within 42 days of delivery; and
- Equity of care indices: child utilization of outpatient services; adult utilization of outpatient services; adult morbidity; and institutional delivery.

Sample design

The sample for this evaluation is a multi-stage probability sample of the selected nine provinces, due to insecurity in two provinces of Kandahar and Kunduz, which were excluded at baseline in 2010-2011. In order to maintain consistency with the baseline RBF assessment, the end-line survey in 2015 was also not conducted in Kandahar and Kunduz. The sample for the baseline sample was a multi-stage probability sample covering populations of the selected nine provinces. In the first stage of sampling health facilities in the nine provinces had been stratified by facility type and the required number of matched health facilities was randomly selected. Health facilities were matched by health facility type and number of outpatient visits in the previous year. In the second stage, the required number of villages or clusters was randomly sampled from the list of all villages in the catchment area of the selected health facilities (2 clusters per health facility). In the third stage, using the household listing conducted prior to survey, the required number of households in the selected villages was sampled using simple random sampling (11).

The same clusters were used in the post-intervention evaluation as were used in the baseline survey. Household lists were created for the baseline survey and updated in the field in the 2015 endline survey. Simple random sampling was used to select 24 households from each cluster. Revisiting the same households and clusters results in greater statistical efficiency producing estimates with smaller variance compared to completely independent samples at two time points (12).

Initial sample size calculation

The sample size was calculated based on the coverage of pentavalent 3 vaccination for children 12-23 months, with the expectation of finding approximately five 12-23 months old children in 20 households and a level 2 and level 3 variation of 0.067 and 0.011, respectively, a standardized effect size (δ) of 0.2 (20%) and significance level (α) of 0.05, resulting in a sample of 140 total health facilities (70 facilities each in the treatment and comparison group groups) with two villages per health facility for 84% power. This is considered a reasonable balance between power and logistical considerations (this calculation takes into account a 7% non-response). This estimation will net a total of 1,300 12-23 months old in 5,564 households (5,954 including 7% non-response). The initial sample size did not take into account the matching design. Generally, effective matching increases the power of the study, as the between-cluster variation within the pairs is smaller than the variation among all the clusters. These calculations were used for determining sample size at baseline. The same sizes apply to the post-intervention evaluation, is 6,912 households, or 288 clusters with 24 households in each cluster in 9 provinces.[5]

Ex-post power estimates

A follow-up power analysis was performed taking into consideration the actual sample size and variances observed as well as the specific tests employed for assessing our null hypothesis that there are no differences between treatment groups for each indicator. The detailed analysis is provided in Appendix 1. The sample size ex-post is able to detect an effect size of 0.2 at 99% power for both 2010 and 2015 surveys and 94% power for the difference in difference analysis, compared to the estimated 0.2 at 84% power from the ex-ante estimation. This is expected since we sampled more than the original sample size estimate, with less than the estimated 7% nonresponse and 24 households per cluster.

Questionnaires

The RBF Household Survey consisted of two questionnaires: (i) the Head of Household Questionnaire, and (ii) the Female and Child Health Questionnaire. The model questionnaires designed by the World Bank were adapted for Afghanistan in consultation with key stakeholders, including the Ministry of Public Health.

The Household Questionnaire was used to collect information from the head of the household on all usual members and visitors in the selected household, including name, sex, education, survival of parents for children under age 18, and utilization of health services. Housing characteristics, such type

[5] Reasons for increased samples at data collection are not clear. We returned to the same households identified at baseline through the household lists. The sample was predetermined at endline.

of water source, sanitation facilities, quality of flooring, and ownership of durable goods, were also collected through the Household Questionnaire and used to determine socio-economic status.

The Female and Child Health Questionnaire was used to collect information from ever-married women age 15 to 49 or women age 18 and over who are primary caretakers of children under 5 years of age on the following topics:

- Background characteristics, including age, literacy and media exposure,
- Pregnancy-based reproductive history,
- Knowledge and use of family planning methods,
- Access, utilization, and quality of services for ANC, delivery, postnatal care, and child health services, including immunization,
- Perception and quality of services and trust in healthcare providers
- Utilization of community-based systems.

These tools are the same as used at baseline and include minor improvements to better facilitate data collection.

Data collection fieldwork

Training for the fieldwork of the endline survey was conducted for two weeks in May 2015. Training materials were similar to the baseline survey, which was conducted in September 2010. Training included lectures on interviewing techniques and the contents of the questionnaires, as well as mapping and listing. Training included lectures on the background of RBF, data quality procedures, how to complete the questionnaires, mock interviews and field work. Throughout the training sessions tests were administered to assess the knowledge and understanding of participants on concepts covered. The tests were used in the selection of surveyors, as well as a feedback mechanism for the trainers for which topics required additional time. Toward the end of the training period, trainees held practice interviews in Kabul with time for feedback from the trainers. For the fieldwork, data were collected by 15 teams. Each team was comprised of six individuals: four interviewers (two males/ two females), one editor, and a male supervisor. Fieldwork for the endline survey began in May 2015 and concluded in July 2015. The baseline survey began in October 2010, and was completed in July 2011. Field teams visited the same clusters and households surveyed during the baseline using information and households maps available from the baseline survey. For clusters where complete household lists were not available (less than 10%), households were randomly selected.

A total of 12 Regional Based Coordinators and Monitors (both male and female) and 2 Territory Field Managers were employed in order to monitor the quality of the data collection process and to facilitate data movement to the Kabul office. Short-refresher courses were conducted as needed and additional monitoring visits were conducted by the MoPH officials. The monitors and coordinators were given training on utilizing monitoring checklists and monitoring/auditing forms to ascertain the performance of the teams under their supervision. Post-monitoring of questionnaires was conducted for 5-10 percent of households.

Data management and analysis

Data processing

All questionnaires were processed at the SRTRO Kabul Office using CSPro, a joint software product of the U.S. Census Bureau, Macro International, and Serpro S.A. The process was ongoing as data were collected. Data processing included data editing, assignment of unique identification numbers for each household, double entry of data, verification of data, and additional data editing that was identified in the double entry or the verification of the data. This was followed by the process of secondary editing which performed consistency checks on the data collected in accordance with guidelines developed for this process. Data processing was completed in August 2015. Data were analyzed using the Stata analysis and statistical software, Version 12 (13).

Coverage of sample and data limitations

The security situation in Afghanistan is constantly fluctuating, making it difficult to predict which health facilities or villages will be accessible at any given time. The roads going into the village or the village itself may be insecure and prevent access for data collectors. Of the originally selected 144 health facilities at baseline, a total of 139 facility catchment areas were sampled in 2015. A total of 13 clusters were dropped due to inaccessibility from insecurity or environmental factors such as flooding or displacement of villages. Tables 1a and 1b present the sample coverages from the 2010 baseline survey and the 2015 endline survey, respectively.

For the baseline survey, of the 7,057 households selected (3,537 and 3,520 selected for treatment and comparison group, respectively) 6,878 interviews were completed (97.5% response), and of those households occupied, 98.1 percent were interviewed. The response rate for eligible women was 97.7 percent. There were no differences in household response rates and eligible women response rates between the comparison group and intervention arms.

Table 1 Results of the household and individual interviews at baseline						
Number of households, number of interviews, and response rates, according to residence (unweighted), 2010						
	Intervention Arm					
	Treatment		Comparison		Total	
	Number	Percent	Number	Percent	Number	Percent
Number of facility catchment areas	72	-	72	-	144	-
Number of clusters	144	-	144	-	288	-
Number of clusters excluded due to insecurity/environmental	0	-	0	-	0	-
Mean households per cluster	24		24			
Household interviews						
Households selected	3,537	100	3,520	100	7,057	100
Households occupied	3,518	99.5	3,491	99.2	7,009	99.3
Households interviewed	3,442	97.3	3,436	97.6	6,878	97.5
Household response rate		97.8		98.4		98.1
Interviews with women age 15--49						
Number of eligible women interviewed	4,405		4,322		8,727	
Eligible women response rate		97.5		97.9		97.7

For the endline survey, of the 6,912 households selected (3,455 and 3,457 selected for treatment and comparison groups, respectively) 6,584 interviews were completed (95.3% response), and of those households occupied, nearly 100 percent were interviewed. Almost equal clusters and health facility catchment areas were dropped between the treatment and comparison groups due to insecurity or environmental factors. However, due to the 13 dropped clusters, the final number of matched-pair health facilities included in the analysis tests is 67. The response rate for eligible women was 98.4 percent. There were no differences in household response rates and eligible women response rates between the comparison and intervention groups. Further descriptions of the dropped clusters are provided in the Annex.

Table 2 Results of the household and individual interviews at endline						
Number of households, number of interviews, and response rates, according to residence (unweighted), 2015						
	Intervention Arm					
	Treatment		Comparison		Total	
	Number	Percent	Number	Percent	Number	Percent
Number of facility catchment areas	70	-	69	-	139	-
Number of clusters	138	-	137	-	275	-
Number of clusters excluded due to insecurity/environmental	6	-	7	-	13	-
Mean households per cluster	24	-	24	-	-	-
Household interviews						
Households selected	3,455	100	3,457	100	6,912	100
Households occupied	3,302	95.6	3,285	95.0	6,587	95.3
Households interviewed	3,300	95.5	3,284	95.0	6,584	95.3
Household response rate		99.9		100		100
Interviews with women age 15-49						
Number of women (15-49 years) interviewed	3,542		3,493		7,035	
Eligible women response rate	3,482	98.3	3,440	98.5	6,922	98.4

At baseline a full list of PSUs or facilities in RBF provinces was obtained from the MOPH health management information system. Once facilities were sampled, the full lists of villages in facility

service areas (i.e. by facility) were obtained from the MOPH HMIS. Clusters were randomly sampled from these lists. The list of names of selected facilities and clusters was shared with MOPH and provincial health officers to ensure their accuracy at the time of the baseline survey.

To ensure accurate sample frame and avoid coverage bias, household mapping and listing was conducted prior to selection of the households. The teams prepared two maps per cluster: (a) first map was for the roads and surrounding areas, and (b) second map was of the village or town itself. Also, during the listing process teams verified village and town names and location. The endline household survey was conducted using the same list of health facilities and catchment areas (i.e. no mapping and listing was conducted).

Data analysis

Analysis of the RBF household survey data was conducted in two phases. The first phase was to replicate the original analysis by JHU based on existing analysis plans and Stata do-files. Details of this analysis are provided in the Annex. The next phase aimed to correct or improve the previous analysis as appropriate. We describe the revised analysis approach.

Calculation of estimates for all indicators was conducted by two independent analysts, using standard definitions from Demographic and Health Survey (DHS) and Multiple Indicator Cluster Survey (MICS). Private health facilities were excluded since the RBF intervention targeted public health facilities only. Table 2 presents the indicator definitions. All data analyses were performed using Stata version 12.0 (13).

Sample weights were calculated for each group to account for selection probabilities. Clustering of observations at the village level, to adjust for within cluster levels of homogeneity (or the intra-cluster correlation), was taken into account. Additionally, strata for application of sampling weights were specified based on matched-pair facility codes since treatment was randomized at this level.

Exploratory data analysis was conducted to check for duplicate household and respondent ID codes, consistency of data between the household and individual questionnaires, presence of missing, unusual and unacceptable values. At baseline, missing values for indicators such as use of modern family planning methods, antenatal care services, skilled birth attendance, or postnatal care was treated as "none" and this was also done at endline. The same methods for developing indicators were used for the endline analyses. In the analyses represented in this reported, missingness was usually less than 1%. Similar approaches were used when developing child health indicators. Descriptive frequencies and statistics were run for characteristics of households and respondents and t-tests or chi-square tests were used to test for differences in groups at baseline and endline.

A wealth index score was constructed for each household using principal component analysis of household assets, sources of income, housing characteristics and water source (14). Principle component analysis generated a weight or factor score for each asset. For each asset, households were assigned a score which was summed by household and individuals within a household were ranked

according to their household score. The sample was divided into quintiles that ranged from one (lowest) to five (highest).

Equity of care indices were estimated by constructing concentration indices and concentration curves. The index was estimated for use of outpatient health services, institutional deliveries, and self-reported morbidity. The calculation followed the procedures described in O'Donnell et al (2008) for micro-data, that is, when values for health variables and previously constructed wealth measures are available for each individual observation (15). The concentration index is a measure of inequality that is estimated as the transformation of a variable of interest on fractional rank of wealth for a population. It ranges from -1 to 1, i.e. from a perfect pro-poor distribution to a pro-rich distribution. The concentration curve displays the share of health variable accounted for by cumulative proportions of individuals in the population ranked from poorest to richest. We tested the dominance of the curves to assess whether the ranking of the treatment and comparison group curves by inequality are unambiguous or statistically significant. We test the study group curves against the 45 degree line of equality and against each other for dominance. Dominance of a curve over another indicates that they are statistically significantly different and the dominant curve is more pro-poor. Non-dominance indicates that there are not differences between the two curves.

For each indicator, household-level matched-pair proportions were tested by treatment group using Chi-square tests and p-values confirmed by logistic regression. Survey design features were declared and Stata was informed of the design variables included by using svy commands. We also used a difference-in-difference (DiD) approach for a panel dataset for estimating the effect of the intervention while accounting for unrelated time trends. The analysis was completed at the household level for matched-pair facilities to compare treatment and comparison groups at baseline and endline. The DiD analysis was conducted for all indicators by using logistic regression models. The DiD estimates by time and study groups are reported in the results. A negative estimate is in favor of the comparison group and a positive estimate is in favor of the intervention group. We also provide a before and after analysis of baseline and endline using logistic regression and tested with an adjusted Wald test.

Proportions and their standard errors are reported for all indicators with their related baseline or endline p-value, DiD estimates, and test p-value. Concentration indices and their confidence intervals and concentration curves are presented for the equity indicators.

Table 3 Indicator definitions

Indicator	Definition
CPR	Proportion of married, non-pregnant women of reproductive age who or whose partner are currently using at least 1 modern family planning method
Skilled ANC	Proportion of women who attended at least 1 ANC visit with a skilled provider of their most recent delivery
SBA	Proportion of births attended by a skilled attendant for a woman's most recent delivery
Pentavalent 3	Proportion of children 12-23 months of age getting three doses third doses of DPT/PENTA before their 1st birthday
Postnatal care (PNC)	Percentage of women who received at least one postnatal care from a trained provider within 42 days of delivery
Equity of care	Is an index where zero indicates equal utilization by all wealth groups; negative numbers indicate higher rates of utilization by poorer patients, while positive numbers indicate higher rates of utilization by wealthier patients
	Institutional deliveries: Proportion of women who had a live birth within the two years preceding the survey and delivered in a public health facility
	Utilization of outpatient services by children under five years: Proportion of children under five years of age reported visiting a public clinic or hospital in the past month and not staying overnight
	Utilization of outpatient services by children over five years and adults: Proportion of children over five years of age and adults reported visiting a public clinic or hospital in the past month and not staying overnight
	Adult morbidity: Proportion of children over five years of age and adults reporting an illness in the two weeks preceding the survey

Results

The results are organized as follows: 1) household and resident characteristics, and 2) health indicators, equity indicators, and before and after analysis.

Household and resident characteristics

Table 2 shows household and resident characteristics at baseline and endline. At end-line, we have 67 matched-pair health facilities. Because we included the survey design in the analysis, it is possible to have matched-paired analysis at the household level, which would appropriately reflect the power of the study and the test p-values. At endline, we have 6,523 matched-pair households.

Overall, absolute differences between treatment and comparison groups at both surveys time points are small. The percent of ever educated woman and wealth distribution are statistically significantly different between the treatment and comparison groups at baseline. There is an increase in the percentage of ever educated women from baseline and the difference is no longer significant at endline. The two groups continue to be different by wealth status at endline. The mean age of women has also slightly increased at endline for both groups. The mean number of residents per household has slightly declined. The distribution of the number of eligible children and ever-married women of reproductive age (15-49) similar across groups.

Table 4 Selected household and resident characteristics*

	Baseline 2010			Endline 2015		
	Treatment	Comparison	p-value	Treatment	Comparison	p-value
Mean number of residents per household (n)	8.1	7.9	0.071	7.3	7.2	0.132
Children 5 years or more and adults (%)	50.5	49.5	0.754	50.5	49.5	0.291
Percent of ever educated women (%)	7.0%	5.4%	<0.01	8.9%	9.7%	0.635
Mean age of women (years [SD])	31.4 (9.2)	31.4 (9.1)	0.946	32.6 (9.0)	32.7 (9.1)	0.554
Ever-married women of reproductive age (15-49) interviewed (%)	50.3	49.7	0.017	50.4	49.6	0.627
Ever-married women of reproductive age with a live birth <24 months prior to survey (%)	50.2	49.8	0.782	50.9	49.2	0.329
Number of eligible children (<5 years of age) (%)	50.5	49.5	0.687	51.2	48.8	0.142
Children 12-23 months (n)	50.9	49.1	0.756	50.1	49.9	0.734
Travel time nearest health facility (< 2 hours) (%)	48.1	51.9	0.279	51.3	48.7	0.812
Wealth quintile			<0.05			<0.01
Lowest	20.7	19.4		17.5	22.3	
Second	18.1	21.9		18.0	21.9	
Middle	18.6	21.4		21.0	19.1	
Fourth	20.0	20.1		20.9	19.0	
Highest	22.7	17.3		22.6	17.6	

*p-values derived from t-tests or chi-square tests

RBF Indicators on maternal and child health and equity

Both study groups made improvements on all RBF health indicators since baseline. At endline, the treatment group shows higher coverage in all indicators compared to the comparison group except for modern contraceptive use. Large improvements are particularly seen for coverage of PENTA 3 vaccination and skilled birth attendance. While PENTA3 vaccination coverage is higher in the

treatment group at endline, the DiD estimate is only 0.2. The highest DiD percentage point change in favor of the treatment group is seen for post-natal care checkup at 7.0 percentage points difference, but this is not statistically significant. No statistically significant differences were found between intervention and comparison groups at endline. The DiD analysis also did not find any differences that were statistically significant.

Table 5 RBF indicators health indicators

Indicator	Baseline			Endline			DiD**	
	Intervention Proportion (SE)	Comparison Proportion (SE)	p-value*	Intervention Proportion (SE)	Comparison Proportion (SE)	p-value*	Diff.	p-value
Current use of modern contraceptives	9.5 (0.8)	9.8 (0.8)	0.770	13.5 (1.1)	16.1 (1.5)	0.127	-2.0	0.394
At least one antenatal checkup from a skilled provider	56.3 (3.1)	61.8 (3.2)	0.155	74.0 (2.7)	71.7 (2.4)	0.508	4.8	0.353
Skilled birth attendant present at last delivery	27.5 (2.2)	23.8 (2.2)	0.220	49.5 (3.1)	42.3 (3.8)	0.108	1.7	0.949
Postnatal checkup within 42 days of delivery by a skilled provider	25.1 (2.4)	25.5 (2.1)	0.879	44.5 (3.6)	37.0 (3.3)	0.114	7.0	0.205
Children (12-23 months) received Pentavalent 3 vaccination	34.7 (3.1)	29.1 (2.9)	0.205	77.5 (3.2)	73.0 (3.4)	0.318	0.2	0.876

*Chi-square test of proportions
**Estimates from panel datasets

The equity indicators presented in Table 6 show the proportions and p-values with the DiD test on the proportions in the first row. Child utilization of outpatient services is statistically significantly different at baseline and endline with the comparison group having higher utilization rates. However, the DiD analysis shows that the 4.3 percentage point difference in favor of the comparison group is not statistically significant. On the other hand, adult utilization of outpatient services has a 7.5 percentage point difference and this is statistically significant, in favor of the comparison group. Utilization rates for the intervention group did not change since baseline. Adult morbidity is higher in the intervention group, meaning more households in the intervention group are self-reporting illness compared to the comparison group, by 4 percentage points and this is statistically significant.

The concentration indices are presented in the second row with their associated confidence interval. At baseline, the treatment group was more pro-poor in the utilization of public health facilities for outpatient care services for children and adults, and by 2015, became less pro-poor. The treatment group was less pro-poor at baseline for institutional deliveries in public health facilities and remains less pro-poor compared to the comparison group at endline. At both baseline and endline, treatment groups have greater morbidity among the poor.

We tested for dominance of the concentration curves (CC), which are presented in Figures 1 for all four equity indicators.[6] The decision rule used here is the multiple comparison approach (mca) with 19 comparison points. With this rule, we reject the null of nondominance in favor of dominance one curve if it is significantly dominant over the other at least at one point and there are no significant

[6] Only endline curves are presented here. Annex 4 presents baseline curves.

differences in the rest of the comparison points.[7] The first test was for dominance between the study group CC and the line of equality at 45 degrees (45DL) to assess whether the respective CC is pro-poor or pro-rich. The second test of dominance tested whether one of the treatment or comparison group CC was statistically significantly dominant over the other. If there is no difference, then the test result is indicated as non-dominant (ND). If either study group or the 45DL was dominant (D), then they are identified as such by the D under their column heading or under the p-value column. At endline, the only statistically significant difference in CCs is adult morbidity, where the treatment CC is dominant – more "pro-poor" or more prevalent among the poor – than the comparison group CC. There are no statistically significant differences in equity based on their CC across the remaining three indicators.

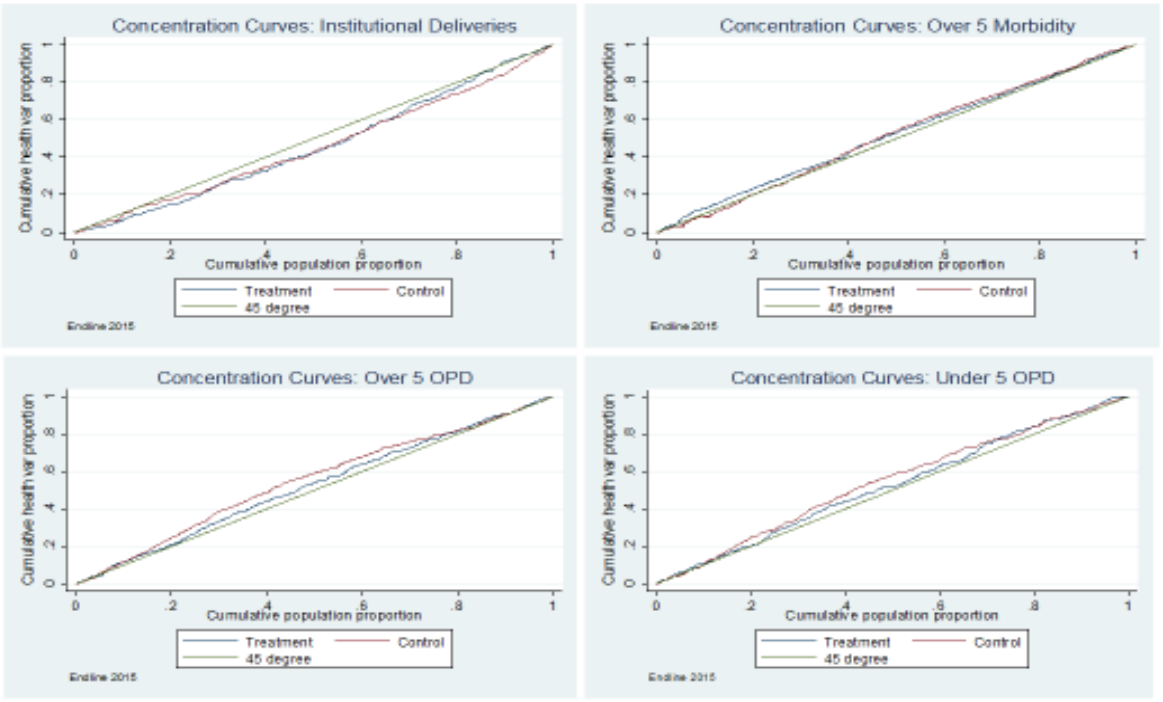
Table 6 RBF equity indicators

Indicator	**Baseline** Intervention Proportion (SE) Equity score (CI)	Comparison Proportion (SE) Equity score (CI)	p-value*	**Endline** Intervention Proportion (SE) Equity score (CI)	Comparison Proportion (SE) Equity score (CI)	p-value*	**DiD**** Diff.	p-value (mean)
Institutional delivery	21.4 (2.3) 0.128 (45DL) (0.022,0.233)	19.7 (2.4) 0.120 (45DL) (0.032,0.208)	0.595 ND	41.0 (0.026) 0.097 (45DL) (0.022, 0.172)	38.4 (0.037) 0.080 (ND) (0.003, 0.157)	0.535 ND	-0.7	0.695
Child (< 5 years) utilization of outpatient services	37.4 (3.8) -0.037 (ND) (-0.144,0.071)	46.1 (2.9) 0.006 (ND) (-0.062,0.073)	<0.10 Trt	32.3 (3.2) -0.061 (ND) (-0.137,0.014)	43.8 (2.7) -0.085 (D) (-0.148,-0.022)	<0.01 ND	-4.3	0.369
Adult (> 5 years) morbidity	24.0 (1.3) -0.029 (D) (-0.076,0.017)	25.8 (1.3) 0.024 (D) (-0.014,0.062)	0.246 Trt	13.1 (0.74) -0.033 (D) (-0.081,0.014)	11.6 (0.7) -0.030 (45DL) (-0.083,0.023)	0.117 Trt	4.0	<0.05
Adult (> 5 years) utilization of outpatient services	31.3 (2.8) -0.064 (D) (-0.146,0.019)	35.3 (2.2) -0.042 (D) (-0.103,0.022)	0.243 ND	31.3 (3.5) -0.045 (ND) (-0.146,0.056)	41.3 (3.1) -0.082 (D) (-0.146,-0.018)	<0.05 ND	-7.5	<0.10

*Chi-square test of proportions; **Estimates from panel datasets; Trt=Treatment group; ND=non-dominant

[7] For more on this method and the related Stata codes see O'Donnell O et al. (2007) chapter 7.

Figure 1 Concentration Curves for Equity Indicators of Utilization



We conducted a before and after analysis by removing the treatment design to test the improvements over time. We cannot attribute any changes to the intervention and this analysis does not show causality. Utilization of outpatient services for children and adults are not statistically significant in this analysis, while the DiD above shows that there is a difference for adult utilization of services. Although reported adult morbidity has declined since baseline, there are no differences in the utilization of outpatient services. This may indicate an inappropriate use of outpatient services. Child utilization of outpatient services declined since baseline, which can be due to less sick children or changes in care seeking practices. On the other hand, while adult morbidity declined since baseline, adult utilization of outpatient services increased. The largest improvement is in the coverage of PENTA 3 vaccination by an absolute increase of 43.3 percentage points.

Table 7 RBF indicators, before and after test*

Indicator	2010 Before	2015 After	Difference	p-value
	Proportion (SE)	Proportion (SE)		
Current use of modern contraceptives	9.6 (0.8)	14.8 (1.0)	5.2	<0.01
At least one antenatal checkup from a skilled provider	60.1 (2.7)	72.7 (2.1)	12.6	<0.05
Skilled birth attendant present at last delivery	26.8 (1.7)	46.2 (2.8)	19.4	<0.01
Postnatal checkup within 42 days of delivery by a skilled provider	26.4 (1.8)	40.9 (2.7)	14.5	<0.01
Children (12-23 months) received Pentavalent 3 vaccination	32.1 (2.2)	75.4 (2.7)	43.3	<0.01
Institutional delivery	21.5 (1.9)	39.8 (2.5)	18.3	<0.01
Child (< 5 years) utilization of outpatient services	42.1 (2.9)	37.7 (2.3)	-4.4	0.623
Adult (> 5 years) morbidity	24.7 (1.1)	12.4 (0.6)	-12.3	<0.01
Adult (> 5 years) utilization of outpatient services	34.2 (2.2)	36.4 (2.9)	2.2	0.103

*Estimates from panel datasets

Descriptive Tables

We present descriptive tables of the main indicators by age of women, education levels (mothers or head of household), travel time to health facility, wealth status, and study group.

Table 8 shows the current use of modern contraceptives by background characteristics of the women. The majority of women who use contraceptives had never attended school. Women from wealthier households were more likely to use contraceptives compared to women from poorer households. Most women lived less than two hours from the nearest health facility, yet among those who lived more than two hours, 13.3% used contraceptives. By age group, women in the younger age groups were less likely to use contraceptives as well as women in older age groups. Women between 25 to 40 years had higher reported uses.

Table 8 Current use of modern contraceptives: Percent distribution of currently married women who use modern contraception methods (N=6,293), weighted %

Background characteristics	n	%	95% CI
Total	886	14.8	(12.9, 16.9)
Current age (years)			
<20	12	3.5	(1.8, 6.7)
20-24	46	3.4	(2.3, 5.0)
25-29	148	12.4	(9.8, 15.6)
30-34	153	17.3	(13.9, 21.3)
35-39	217	22.6	(18.6, 27.2)
40-44	181	21.9	(18.1, 26.3)
45-49	129	16	(12.7, 20.1)
Education			
Never attended school	813	15.2	(13.2, 17.3)
Primary school	27	10.2	(5.6, 17.6)
Secondary school and more	43	11.6	(7.9, 16.7)
Travel time to health facility			
Nearest health facility < 2 hours	474	14.9	(12.7, 17.4)
Nearest health facility > 2 hours	58	13.3	(9.3, 18.6)
Wealth quintile			
Lowest	153	12.7	(10.1, 15.8)
Second	175	14	(10.9, 17.7)
Middle	182	15.2	(12.4, 18.5)
Fourth	168	16.1	(12.6, 20.3)
Highest	181	15.8	(12.7, 19.6)
Intervention			
Comparison	469	16.1	(13.4, 19.3)
Treatment	417	13.5	(11.4, 16.0)

Table 9 shows the distribution of antenatal care coverage by background characteristics. About 73% of women had at least one ANC visit to a skilled provider for the most recent live birth in the two years preceding the survey. Among these, women in the middle ages of childbearing were more likely to access ANC as well as women who had some primary or secondary education. Women who lived closer to a health facility had greater use of ANC (78.3%) compared to women who lived further away (58.9%). Women from wealthier households were also more likely to access ANC.

Table 9 Antenatal care coverage: Percent distribution of ever-married women who had at least one ANC visit to a skilled provider for the most recent live birth in the two years preceding the survey (excluding currently pregnant women) (N=2,085), weighted %

Background characteristics	n	%	95% CI
Total	1516	73.0	(68.8, 76.8)
Current age (years)			
<20	77	68.9	(56.1, 79.3)
20-24	379	76.5	(70.5, 81.5)
25-29	388	73.8	(68.0, 79.0)
30-34	291	73.5	(67.0, 79.0)
35-39	235	69.9	(60.7, 77.7)
40-44	111	68.5	(60.1, 75.8)
45-49	35	68.1	(48.6, 82.8)
Education			
Never attended school	1317	71.8	(67.7, 75.5)
Primary school	72	79.9	(62.5, 90.5)
Secondary school and more	122	82.0	(69.7, 90.1)
Travel time to health facility			
Nearest health facility < 2 hours	906	78.3	(72.6, 83.1)
Nearest health facility > 2 hours	78	58.9	(48.5, 68.6)
Wealth quintile			
Lowest	221	63.6	(56.0, 70.7)
Second	286	72.6	(65.4, 78.9)
Middle	280	77.2	(70.9, 82.5)
Fourth	309	27.3	(65.5, 79.0)
Highest	379	78.4	(71.9, 83.7)
Intervention			
Comparison	731	71.7	(66.6, 76.3)
Treatment	785	74.0	(68.2, 79.0)

Skilled birth attendance by background characteristics is presented in Table 10. Younger women (<30 years) were more likely to use skilled attendants at delivery. Women who had secondary level education or more had much higher use of SBA (65.4%). Women who lived closer to a health facility and who were from wealthier households are also more likely to use SBA.

Table 10 Percent distribution of skilled birth attendance: Use of skilled attendants by respondent characteristics (N=2,268), weighted %

Background characteristics	n	%	95% CI
Total	1061	46.3	(41.1, 51.7)
Current age (years)			
<20	61	51.9	(38.1, 65.4)
20-24	307	55.5	(47.6, 63.1)
25-29	262	43.5	(36.9, 50.3)
30-34	175	39.8	(33.2, 46.9)
35-39	162	43.3	(33.6, 53.6)
40-44	73	43.8	(33.6, 54.5)
45-49	21	38.0	(22.6, 56.2)
Education			
Never attended school	899	44.6	(39.4, 49.9)
Primary school	59	49.3	(37.6, 61.1)
Secondary school and more	99	65.4	(52.5, 76.4)
Travel time to health facility			
Nearest health facility < 2 hours	675	52.4	(46.2, 58.5)
Nearest health facility > 2 hours	46	32.8	(22.6, 45.0)
Wealth quintile			
Lowest	124	35.9	(29.0, 43.4)
Second	170	38.5	(30.9, 46.6)
Middle	180	41.6	(33.4, 50.3)
Fourth	234	48.8	(41.5, 56.2)
Highest	325	63.2	(55.7, 70.0)
Intervention			
Comparison	478	42.3	(34.9, 50.0)
Treatment	583	49.5	(43.3, 55.8)

The percent distribution of postnatal care by a skilled provider is presented in Table 11. Similar distributions by characteristics for SBA are seen for PNC. Younger women, women with higher levels of education, women who live closer to the health facility, and women from wealthier households are more likely to use PNC services.

Table 11Percent distribution of postnatal care received by a skilled provider within 42 days of delivery by respondent characteristics (N=2,268), weighted %

Background characteristics	n	%	95% CI
Total	915	41.2	(36.0, 46.6)
Current age (years)			
<20	58	50.4	(36.9, 63.9)
20-24	254	49.6	(40.5, 58.6)
25-29	217	38.2	(32.0, 44.8)
30-34	153	32.8	(26.5, 39.8)
35-39	148	41.2	(30.6, 52.7)
40-44	67	39.1	(30.0, 49.1)
45-49	18	27.7	(15.4, 44.6)
Education			
Never attended school	783	39.9	(34.9, 45.1)
Primary school	45	44.3	(29.7, 60.0)
Secondary school and more	83	54.6	(43.0, 65.7)
Travel time to health facility			
Nearest health facility < 2 hours	574	45.8	(38.8, 52.9)
Nearest health facility > 2 hours	47	31.9	(21.0, 45.2)
Wealth quintile			
Lowest	119	36.0	(29.9, 42.7)
Second	141	30.4	(24.7, 36.8)
Middle	159	37.4	(28.7, 47.0)
Fourth	203	45.5	(37.3, 54.0)
Highest	271	54.1	(48.4, 59.7)
Intervention			
Comparison	410	37.0	(30.7, 43.9)
Treatment	505	44.5	(37.5, 51.7)

Overall institutional deliveries remains low in the country (39.9%). Younger women are more likely to deliver in a health facility. Over 50% of women with secondary level education or more delivered in a health facility compared to 38.6% of women with no education. Women from wealthier households are also more likely to deliver in a health facility.

Table 12 Institutional deliveries: Percent distribution of ever-married women who had a live birth in the two years preceding the survey with an institutional delivery for their most recent birth (N=2,268), weighted %

Background characteristics	n	%	95% CI
Total	905	39.9	(35.3, 44.7)
Current age (years)			
<20	46	45.2	(31.8, 59.4)
20-24	274	49.9	(42.7, 57.1)
25-29	219	38.0	(32.2, 44.2)
30-34	149	32.5	(27.5, 38.0)
35-39	133	35.2	(26.7, 44.7)
40-44	68	38.1	(27.4, 50.2)
45-49	50	26.3	(14.4, 43.1)
Education			
Never attended school	775	38.6	(34.0, 43.5)
Primary school	51	42.8	(29.9, 56.7)
Secondary school and more	76	52.2	(41.3, 62.9)
Travel time to health facility			
Nearest health facility < 2 hours	577	44.9	(39.8, 50.1)
Nearest health facility > 2 hours	41	26.6	(16.4, 40.1)
Wealth quintile			
Lowest	114	32.1	(24.4, 41.0)
Second	161	35.2	(28.6, 42.4)
Middle	162	37.1	(29.0, 45.9)
Fourth	203	43.9	(37.6, 50.3)
Highest	239	48.0	(39.6, 56.5)
Intervention			
Comparison	420	38.4	(31.3, 46.1)
Treatment	485	41.0	(35.9, 46.3)

Table 13 shows the distribution of children ages 12 to 23 months who received PENTA3 vaccination by their background characteristics. There were no difference in vaccination coverage by sex. Younger mothers were more likely to have their children vaccinated compared to older mothers. At least primary school education of the mother showed higher vaccination coverage.

Table 13 Percent distribution of children 12-23 months who received PENTA3 vaccinations at any time before the survey by health card or mother's report, (N=1,038), weighted %.

Background characteristics	n	%	95% CI
Total	771	75.5	(70.1, 80.2)
Child's sex			
Female	365	75.0	(68.0, 80.9)
Male	404	76.0	(69.5, 81.5)
Mother's current age (years)			
<20	37	79.8	(62.9, 90.2)
20-24	189	74.9	(68.8, 80.2)
25-29	195	77.0	(68.8, 83.5)
30-34	138	75.7	(66.4, 83.1)
35-39	136	76.7	(63.9, 86.0)
40-44	53	70.4	(53.3, 83.2)
45-49	23	63.5	(40.8, 81.5)
Mother's education level			
Never attended school	685	75.4	(69.7, 80.3)
Primary school	31	80.3	(62.7, 90.8)
Secondary school and more	51	76.1	(62.7, 85.7)
Travel time to health facility			
Nearest health facility < 2 hours	459	80.2	(74.3, 85.0)
Nearest health facility > 2 hours	44	73.1	(56.6, 85.0)
Wealth quintile			
Lowest	122	81.2	(71.2, 88.3)
Second	139	67.8	(58.7, 75.8)
Middle	139	73.7	(60.5, 83.7)
Fourth	175	83.0	(74.4, 89.1)
Highest	175	72.2	(64.5, 78.8)
Intervention			
Comparison	378	73.0	(65.7, 79.3)
Treatment	393	77.5	(70.5, 83.2)

Table 14 presents the distribution of children under five years of age who were sick two weeks before the survey. About 43.7% of children reported having been sick, with little difference by child's sex. Children living in households further away from health facilities had higher percentage of illness (52.3%) compared to children who lived closer to the health facility (40.1%).

Table 14 Percent distribution of children under five years of age who were sick two weeks before the survey, (N=5,935), weighted %.

Background characteristics	n	%	95% CI
Total	2633	43.7	(40.3, 47.2)
Child's sex			
Female	1283	44.5	(40.4, 48.7)
Male	1341	42.9	(39.4, 46.4)
Mother's current age (years)			
<20	85	48.0	(38.4, 57.7)
20-24	552	44.5	(39.2, 50.0)
25-29	706	42.5	(38.0, 47.2)
30-34	511	43.1	(38.9, 47.5)
35-39	459	45.3	(40.6, 50.2)
40-44	230	41.6	(34.5, 49.0)
45-49	90	44.0	(33.1, 55.6)
Mother's education level			
Never attended school	2343	43.2	(39.7, 46.8)
Primary school	127	52.8	(46.2, 59.2)
Secondary school and more	156	44.2	(36.4, 52.4)
Travel time to health facility			
Nearest health facility < 2 hours	1721	40.1	(33.3, 47.3)
Nearest health facility > 2 hours	167	52.3	(49.1, 55.6)
Wealth quintile			
Lowest	443	38.5	(32.7, 44.6)
Second	501	44.2	(39.4, 49.1)
Middle	506	46.3	(40.6, 51.9)
Fourth	502	43.7	(37.3, 50.2)
Highest	629	46.6	(42.1, 51.3)
Intervention			
Comparison	1291	40.6	(36.4, 44.9)
Treatment	1342	46.3	(42.2, 50.5)

Of those children who were sick, only 37.2% sought outpatient care services as shown in Table 15. Children from poorer households were more likely to seek health care services compared to children from wealthier households. Children with mothers who had never attended school were also more likely to seek care compared to children with mothers who had secondary education or more. This may be an indication that wealthier households with educated mothers are better able to manage the illness of their sick children.

Table 15 Percent distribution of children under five years of age who were sick two weeks before the survey and sought outpatient care services, (N=2,633), weighted %.

Background characteristics	n	%	95% CI
Total	1032	37.2	(32.5, 42.2)
Child's sex			
Female	491	36.5	(30.3, 43.0)
Male	540	38.1	(33.9, 42.5)
Mother's current age (years)			
<20	34	40.2	(26.6, 55.4)
20-24	211	35.1	(28.3, 42.5)
25-29	263	35.2	(29.5, 41.5)
30-34	203	38.9	(31.6, 46.8)
35-39	185	39.3	(31.4, 47.8)
40-44	98	40.7	(31.5, 50.5)
45-49	38	33.9	(23.5, 46.1)
Mother's education level			
Never attended school	924	37.4	(32.6, 42.4)
Primary school	53	35.8	(22.1, 52.2)
Secondary school and more	49	32.9	(24.2, 42.8)
Travel time to health facility			
Nearest health facility < 2 hours	819	44.3	(38.0, 50.8)
Nearest health facility > 2 hours	69	39.3	(29.3, 50.4)
Wealth quintile			
Lowest	205	44.2	(36.3, 52.3)
Second	210	44.2	(37.4, 51.2)
Middle	188	34.7	(27.0, 43.3)
Fourth	197	37.8	(30.8, 45.4)
Highest	211	28.9	(22.5, 36.4)
Intervention			
Comparison	536	43.8	(38.5, 49.3)
Treatment	496	32.3	(26.4, 39.0)

Tables 16 and 17 present older children and adult morbidity and care seeking by background household characteristics. About 12.4% reported being sick two weeks before the survey. There were no large differences by background characteristic. Of those who reported being sick, about 35.8% reported seeking outpatient care services. Similar to children care seeking practices, adults and older children from less educated and poorer households also sought more care.

Table 16 Percent distribution of children over five years of age and adults who were sick two weeks before the survey, (N=39,451), weighted %.

Background characteristics	n	%	95% CI
Total	4855	12.4	(11.2, 13.6)
Head of household education level			
Never attended school	3702	12.1	(11.1, 13.3)
Primary school	546	15.8	(13.9, 18.0)
Secondary school and more	589	11.6	(9.4, 14.3)
Travel time to health facility			
Nearest health facility < 2 hours	3697	19.7	(17.6, 22.0)
Nearest health facility > 2 hours	552	18.8	(17.6, 20.1)
Wealth quintile			
Lowest	943	13.5	(11.5, 15.7)
Second	972	12.7	(11.1, 14.5)
Middle	942	12.6	(10.8, 14.7)
Fourth	866	11.2	(9.6, 13.0)
Highest	986	11.6	(10.0, 13.3)
Intervention			
Comparison	2392	11.6	(10.2, 13.1)
Treatment	2463	13.1	(11.6, 14.6)

Table 17 Percent distribution of children over five years of age and adults who were sick two weeks before the survey and sought outpatient care services, (N=4,855), weighted %

Background characteristics	n	%	95% CI
Total	1894	35.8	(30.4, 41.6)
Head of household education level			
Never attended school	1472	37.3	(31.7, 43.3)
Primary school	209	32.8	(24.6, 42.3)
Secondary school and more	200	29.2	(21.9, 37.8)
Travel time to health facility			
Nearest health facility < 2 hours	1710	42.7	(35.5, 50.1)
Nearest health facility > 2 hours	184	31.7	(25.5, 38.7)
Wealth quintile			
Lowest	423	41.7	(32.2, 51.9)
Second	406	42.4	(34.4, 50.8)
Middle	352	33.5	(26.3, 41.6)
Fourth	306	28.1	(21.7, 35.6)
Highest	341	31.8	(23.6, 41.2)
Intervention			
Comparison	988	41.3	(35.3, 47.5)
Treatment	906	31.3	(24.7, 38.7)

CHAPTER 2: COMMUNITY LEADERS' SURVEY

A survey of community leaders was conducted at baseline to assess their perceptions of the community's demographic characteristics, economic activities, crises and other shocks, and social capital. This survey was implemented again as part of the follow-up household survey for the RBF impact evaluation.

Methodology

At the time of the RBF endline household survey, the Community Leaders' survey was conducted in the catchment areas of selected RBF facilities in the nine provinces. The survey targeted Community Development Council (CDC) chairpersons or deputy chairpersons of the Shura in the village.

Sample

Sampling of village clusters and leaders followed the design of the RBF household survey. For details of the sampling design and selection, please refer to the household survey report. Leaders were interviewed from selected clusters in the catchment areas of selected RBF health facilities. At endline, leaders from 262 villages were interviewed, 130 from the intervention group and 132 from the comparison group.

Table 18 Sample of community leaders at baseline and endline

	Baseline (2010)			Endline (2015)		
	Intervention	Comparison	Total	Intervention	Comparison	Total
Clusters/Villages	131	126	257	130	132	262
Matched-pair health facilities	62	62	124	70	69	139

Data collection and management

Interviews of leaders was done at the same time as the household survey. Training for the household survey and community leaders' survey was done at the same time over a period of two weeks in May 2015. Training materials were similar to the baseline survey, which was conducted in September 2010. For more information on data collection and fieldwork, please refer to the household survey report. Data processing also occurred at the same time as the household survey by SRTRO Kabul Office using CSPro.

Analysis

The community leaders' survey questionnaire collected data from the following content areas: 1) background, 2) direct observation, 3) demography, 4) basic services, 5) social capital, 6) economic activities, and 7) external shocks. Household survey sample weights and survey design setting were used up to the cluster level strata.

Exploratory data analysis was conducted to check for duplicate cluster ID codes, consistency of data, levels of missing data, and unusual and unacceptable values. Most indicators had less than 1% missing values. The data were survey set according to the survey design and all indicators were analyzed by matched-pair analysis at the cluster level. Descriptive statistics were run for characteristics of community leaders who were interviewed at baseline and endline. Means or proportions and 95% confidence intervals are presented for endline results. Differences in proportions or means were tested using a logistic or linear regression followed by an adjusted Wald test.

Results

Table 1 presents background characteristics of participants. There are no differences at baseline or at endline across the two study groups in the number of months of person serving as leader, age, gender and current job. At baseline, there is a slight statistically significant difference in the age of respondents between intervention and comparison groups but this is likely due to random chance. The average months spent as the chairperson at baseline was about 33 months and at endline, around 40 months. Almost all of the community leaders all male in both groups, with over a third of their current jobs as being a farmer.

Table 19 Background characteristics of participants

Characteristics	Baseline (n=257)			Endline (n=262)		
	Intervention	Comparison	p-value	Intervention	Comparison	p-value
Time as chair (months) (mean, 95% CI)	33.7 (27.0, 40.4)	33.4 (28.4, 38.4)	0.946	39.2 (34.0, 44.4)	44.7 (36.8, 52.6)	0.242
Age (years) (mean, 95% CI)	47.3 (44.6, 50.0)	44.6 (42.7, 46.6)	<0.10	48.8 (46.1, 51.6)	46.5 (44.6, 48.5)	0.184
Gender (%, 95% CI)			0.236			0.535
Male	98.1 (91.5, 99.6)	100		98.3 (93.1, 99.6)	99.2 (94.4, 99.9)	
Female	1.9 (0.4, 8.5)	0		1.7 (0.4, 6.9)	0.8 (0.1, 5.6)	
Current job (%, 95% CI)			0.992			0.208
Farmer	44.9 (33.7, 56.7)	49.9 (37.5, 62.2)		49.0 (36.4, 61.8)	39.3 (28.4, 51.4)	
Teacher	15.5 (9.1, 25.3)	13.9 (7.1, 25.2)		15.2 (7.3, 29.0)	16.5 (9.3, 27.6)	
Elected representative	15.9 (7.3, 31.4)	10.2 (4.6, 21.1)		12.6 (7.3, 21.1)	11.4 (5.9, 20.8)	
Government official	9.9 (5.0, 18.8)	6.9 (3.0, 15.0)		3.0 (0.8, 10.9)	2.7 (0.8, 8.3)	
Other	13.7 (8.2, 21.9)	19.2 (11.6, 30.1)		20.1 (11.6, 32.8)	30.2 (20.1, 42.6)	

Table 2 shows the results of direct observations of the community made by data collectors. There are no differences between the two groups across all direct observations such as children and adults wearing shoes, houses with glass windows, visible garbage, and perceptions of cleanliness of the villages. The intervention and comparison group villages were also similar in topography. An overwhelming majority of the houses in both study groups are made of clay.

Table 20 Direct observations (N= 262)

Observation	Endline 2015				p-value
	Intervention		Comparison		
	n	% (95% CI)	n	% (95% CI)	
Typically wear shoes (children)	97	81.1 (72.3, 87.6)	103	76.9 (67.1, 84.5)	0.413
Typically wear shoes (adults)	120	92.1 (84.4, 96.2)	117	87.4 (79.2, 92.7)	0.302
House windows with glass	113	88.6 (80.3, 93.6)	116	87.7 (78.1, 93.4)	0.858
Cleanliness of house surroundings	89	66.3 (55.1, 75.9)	88	70.4 (60.7, 78.6)	0.504
Visible garbage along roads	42	34.4 (24.6, 45.6)	45	30.3 (21.6, 40.8)	0.521
Visible animal droppings	80	66.5 (54.6, 76.6)	81	60.3 (48.1, 71.4)	0.407
Obstructed sewers/gutters/canals	24	16.3 (9.8, 25.7)	22	18.3 (11.3, 28.3)	0.735
Stagnant pools of water	20	16.8 (10.0, 26.8)	24	19.1 (12.1, 28.8)	0.701
Cleanliness of village compared to others					0.211
Very clean	17	12.8 (6.8, 22.9)	11	6.3 (2.8, 13.9)	
Clean	89	68.7 (58.1, 77.7)	89	71.8 (62.3, 79.8)	
Dirty	23	17.0 (10.4, 26.6)	29	19.5 (13.2, 27.7)	
Very dirty	1	1.5 (0.2, 10.0)	3	2.4 (0.7, 7.6)	
Village topography					0.419
Inland plains	38	28.0 (17.6, 41.4)	40	24.0 (15.9, 34.6)	
Hills	49	39.7 (29.2, 51.2)	44	35.8 (25.2, 47.9)	
Mountains	36	27.9 (18.8, 39.2)	46	39.0 (28.3, 50.8)	
Desert	6	3.9 (1.3, 11.3)	2	1.2 (0.2, 6.7)	
Material of houses					0.797
Bricks	12	11.3 (5.3, 22.4)	11	8.8 (3.7, 19.5)	
Cement	4	0.7 (0.2, 2.6)	0	0	
Clay	98	74.5 (63.8, 82.9)	102	78.6 (68.2, 86.3)	
Other	16	13.5 (7.8, 22.2)	19	12.5 (7.4, 20.4)	

Table 3 shows the average number of households reported by each community leader by study group as well as their population estimates. There are no differences between the two study groups – on average, there are 200 households with an estimated population size of over 1,200 people.

Table 21 Demographic characteristics (N = 262)

Characteristic (mean, 95% CI)	Endline 2015				p-value
	Intervention		Comparison		
	n	% (95% CI)	n	% (95% CI)	
Average number of households	130	202 (170, 234)	132	206 (178, 234)	0.851
Population estimate	130	1315 (1101, 1528)	130	1291 (1133, 1448)	0.854

In terms of access to basic services, Table 4 shows that the intervention group has more tarmac and gravel roads compared to the comparison group, and the comparison group has more dust roads compared to the intervention group. However, no differences were found in the distance to the nearest paved road or reported difficulty in traveling to and from the village. The average distance to the nearest urban center is also very similar, on average 35 km for the intervention group and 37 km for the comparison group. Costs for transportation are also very similar and not statistically significantly different at almost 300 Afs. to the nearest urban center.

Table 22 Basic Services (N=262)

Characteristics	Endline 2015				p-value
	Intervention		Comparison		
	n	% (95% CI)	n	% (95% CI)	
Most common type of road surface					<0.05
Tar/tarmac	7	5.6 (2.4, 12.7)	1	1.3 (0.2, 8.8)	
Gravel	43	27.0 (18.3, 38.0)	29	19.4 (12.4, 29.1)	
Dust	74	65.6 (54.7, 75.1)	98	77.8 (68.1, 85.2)	
Other	4	1.7 (0.5, 5.0)	3	1.5 (0.4, 5.8)	
Distance to nearest paved road (km) (mean, 95% CI)	127	44.8 (27.6, 62.1)	131	38.0 (28.3, 47.7)	0.465
Difficulty traveling to/from village					0.620
Very difficult	41	35.3 (24.9, 47.3)	45	34.8 (24.8, 46.3)	
Difficult	25	22.6 (14.5, 33.5)	32	23.4 (15.6, 33.5)	
A little difficult	29	15.9 (9.6, 25.2)	30	25.5 (15.4, 39.1)	
Easy	34	26.1 (15.9, 39.8)	25	16.4 (10.1, 25.5)	
Distance to nearest urban center (km) (mean, 95%CI)	129	35.5 (20.2, 50.9)	131	37.3 (29.3, 45.4)	0.835
Cost to nearest urban center using public transportations (Afs.) (mean, 95% CI)	130	282 (192, 371)	132	299 (222, 376)	0.755

Table 5 shows the available facilities or services in the village such as primary or secondary schools. There is a statistically significant difference between the intervention (10.5%) and comparison group (3.1%) groups with available government hospitals ($p<0.10$). There is also a statistically significant difference in the availability of private hospitals – 0.1% in the intervention group and 2.6% in the comparison group ($p<0.01$). There are no differences in all other available facilities or services between the study groups.

Table 23 Facilities or services available in village (N = 262)

Facilities/services	Endline				p-value
	Intervention		Comparison		
	n	% (95% CI)	n	% (95% CI)	
Primary school	79	60.7 (58.9, 71.4)	72	53.1 (41.5, 64.4)	0.362
Secondary school	87	71.1 (60.2, 80.0)	91	74.4 (64.4, 82.3)	0.610
Government hospital	14	10.5 (6.8, 15.7)	6	3.1 (0.9, 9.6)	<0.10
Government clinic	87	68.3 (56.7, 78.1)	85	67.5 (57.1, 76.5)	0.909
Private hospital	3	0.1 (0.0, 0.6)	1	2.6 (0.8, 8.7)	<0.01
Private clinic	4	8.7 (2.7, 24.9)	4	2.7 (0.8, 8.7)	0.175
Mosque	118	92.3 (85.0, 96.2)	115	88.9 (80.5, 94.0)	0.425
Pharmacy	30	30.5 (20.4, 43.0)	24	21.2 (12.7, 33.2)	0.241
CDC shura center	64	49.8 (38.3, 61.3)	64	44.9 (34.9, 55.3)	0.506

There are no statistically significant differences in women's rights to purchase, inherit, sell, or leave land between the two study groups.

Table 24 Women's rights (N = 262)

The right to…	Endline				
	Intervention		Comparison		p-value
	n	% (95% CI)	n	% (95% CI)	
Purchase land	28	75.5 (63.9, 84.3)	97	80.7 (71.6, 87.4)	0.423
Inherit land	120	90.5 (81.4, 95.4)	121	94.1 (87.7, 97.2)	0.375
Sell land	110	80.9 (69.3, 88.8)	114	88.2 (80.2, 93.2)	0.202
Leave land	118	88.5 (79.2, 93.9)	122	94.1 (87.7, 97.3)	0.184

Table 7 provides village characteristics related to governance and leadership. There are no differences between the two study groups on governance and leadership characteristics. In the intervention group 86.3% of community leaders are elected and 81.9% are elected in the comparison group. Most decision making occurs with the government and the community – there seems to be little NGO involvement in both groups. Less than three-quarters of communities in both study groups have regular Shura meetings at least once a month. Of those who do have meetings, on average 17 people are present at meetings in the intervention communities and on average 16 people are present in comparison group communities.

Table 25 Village governance and leadership (N = 262)

Characteristics	Endline				
	Intervention		Comparison		p-value
	n	% (95% CI)	n	% (95% CI)	
Selection of village head (%, 95% CI)					0.303
Elected	106	86.3 (76.9, 92.2)	103	81.9 (72.8, 88.4)	
Appointed by elder	20	12.4 (6.7, 21.8)	23	15.0 (9.2, 23.6)	
Appointed by administration	2	0	2	1.4 (0.3, 6.3)	
Participation in decision-making for community development projects (%, 95% CI)					0.863
Government	61	48.2 (34.3, 62.3)	60	51.9 (39.4, 64.2)	
Community	42	51.8 (37.7, 65.7)	45	45.9 (33.3, 60.0)	
NGO microfinance project	0	0	3	2.2 (0.6, 8.1)	
Regular CDC Shura meetings at least once a month (%, 95% CI)	75	64.9 (52.3, 75.8)	83	69.4 (57.9, 78.9)	0.524
Average number of people present at CDC Shura meetings (mean, 95% CI)	102	17.1 (14.2, 19.9)	105	15.8 (14.3, 17.4)	0.409

*CDC=Community Development Council

Table 8 shows village characteristics by economic activity such as sources of livelihood, employment migration, and comparisons of perceived wealth status with other villages and itself. Slightly statistically significant differences ($p<0.10$) are shown for piece work (greater among intervention group) and farming (greater among comparison group) for sources of livelihood. Most villages reported people leaving the village for temporary work in both study groups. But there were statistically significant differences in destinations for work; over 50% of the comparison group go to

urban areas and 28% leave the country whereas 48% of the intervention group go to urban areas and 44% leave the country. More people in the comparison group tend to go to other rural areas as well. Most communities in both study groups perceived themselves poorer than other villages or the same, while more of the intervention group saw themselves poorer than they were 5 years ago (43.6% and comparison group 34.1%). The comparison group considered themselves to be more or less the same (40% and intervention 30.9%).

Table 26 Economic activity (N = 262)

Characteristics	Endline				p-value
	Intervention		Comparison		
	n	% (95% CI)	n	% (95% CI)	
Most important source of livelihood					
Wage employment	43	26.2 (17.6, 37.1)	29	24.7 (15.7, 36.7)	0.835
Piece work	25	21.8 (13.1, 34.0)	15	12.3 (6.7, 21.6)	<0.10
Farming	108	85.6 (71.0, 93.5)	117	95.1 (88.2, 98.1)	<0.10
Other	46	39.5 (28.0, 52.3)	42	37.0 (26.0, 49.5)	0.777
Leave village for temporary work	119	93.1 (85.2, 96.9)	117	86.6 (76.4, 92.8)	0.199
Destination for work outside of village					<0.01
Rural areas	11	7.8 (3.9, 15.1)	13	18.1 (9.7, 31.2)	
Urban areas	68	48.4 (36.3, 60.7)	63	53.7 (41.1, 65.8)	
Outside country	40	43.8 (31.4, 57.0)	41	28.2 (19.1, 39.6)	
Village is poorer-wealthier than other villages					0.115
Poorer	72	55.8 (44.4, 66.6)	62	44.9 (34.2, 56.1)	
The same	51	37.1 (27.8, 47.6)	57	41.1 (30.1, 53.1)	
Wealthier	6	7.1 (2.3, 19.6)	12	13.9 (7.5, 24.5)	
Village is poorer-wealthier than it was 5 years earlier					0.371
Poorer	59	43.6 (32.8, 54.9)	52	34.1 (25.0, 44.6)	
The same	37	30.9 (21.7, 41.9)	45	40.0 (29.5, 51.5)	
Wealthier	33	25.6 (16.9, 36.7)	35	25.8 (17.2, 37.0)	

There were no statistically significant differences between the two study groups in their reported external shocks, except for loss of jobs, which was greater among the intervention group (45.3%) compared to the comparison group (33.2%) ($p<0.10$). The highest reported external shocks by both study groups were drought, flood, and blight. There was little loss of social services or new employment opportunity.

Table 27 External shocks (N = 262)

Event	Endline				p-value
	Intervention		Comparison		
	n	% (95% CI)	n	% (95% CI)	
Drought	85	64.1 (51.6, 75.0)	87	62.0 (50.2, 72.5)	0.756
Flood	75	54.9 (43.1, 66.1)	71	48.9 (37.1, 60.9)	0.474
Blight/Pestilence	81	64.3 (51.7, 75.2)	85	60.5 (47.6, 72.1)	0.618
Livestock disease	72	53.2 (41.3, 64.7)	73	48.8 (37.4, 60.2)	0.576
Human disease epidemic	37	24.7 (17.2, 34.1)	28	19.2 (11.7, 29.9)	0.391
Sharp inflation	52	42.1 (30.7, 54.4)	50	34.1 (24.3, 45.4)	0.213
Massive loss of jobs	56	45.3 (34.1, 56.9)	40	33.2 (23.5, 44.5)	<0.10
Loss of social services	11	7.0 (3.4, 13.7)	9	4.8 (2.2, 10.0)	0.325
New health facility/major improvements to old facility	32	23.5 (14.5, 35.7)	26	19.0 (11.9, 28.8)	0.516
New school / major improvements to old school	35	22.9 (14.0, 35.1)	38	28.7 (18.7, 41.4)	0.448
New road/major improvement to old road	43	29.5 (20.4, 40.7)	34	21.0 (13.5, 31.1)	0.174
New employment opportunity	18	11.5 (6.4, 19.9)	16	11.9 (6.5, 20.7)	0.946

CHAPTER 3: HEALTH FACILITY ASSESSMENT

The health facility assessment was designed to assess the impact of RBF on the performance of health facilities as measured by the Balanced Scorecard (BSC). The BSC provides a framework to efficiently look at several key areas or domains of the health sector.

The domains and indicators of the BSC were developed by the Technical Working Group led by the Ministry of Public Health (MoPH) and involved representatives of the MoPH, implementing agencies, international partners, and health providers. The BSC used for the assessment of RBF consists of the same domains and indicators as the BSC used for the annual National Health Services Performance Assessment (NHSPA), with the exception of Domain F. Domain F consists of two equity indicators – equity of outpatient services and equity of patient satisfaction – that are measured more rigorously through the household survey. There are 20 indicators categorized under the following five domains:

- Domain A: Client and Community
- Domain B: Human Resources
- Domain C: Physical Capacity
- Domain D: Quality of Service Provision
- Domain E: Management Systems

In addition, there is an overall performance score, which is an average of the scores on the 20 indicators from five domains described above. Each indicator is given equal weighting to develop the composite score which ranges from 0 to 100, with better performing facilities receiving higher scores. For more information on the BSC, please see the 2015 final report (16).

Methodology

Sample

Data for the BSC are collected under the NHSPA, which has been conducted annually, using a stratified random sample of health facilities providing the BPHS from each provinces, and systematic random samples of patients and health workers. As a national health facility assessment, all 11 RBF provinces were sampled under the NHSPA.

A sampling frame of BPHS health facilities was created for each province by compiling a list of health facilities from MoPH in Kabul and updating them with the Provincial Public Health Directors (PPHD), HMIS officers and key informants from NGOs and the MoPH in each province. A stratified random sample of 25 facilities providing BPHS services are taken from each province. After finalizing the list of functional facilities, in consultation with the PPHDs, HMIS officers and NGOs, facilities were stratified into three groups: Comprehensive Health Centers (CHC), Basic Health Centers (BHC) and Sub-Health Centers (SHC). For each province, facilities were then randomly selected from each stratum according to the following distribution by facility type: five CHCs, fifteen BHCs and five SHCs. If fewer than the above number of a particular facility type were available for survey in any province, other facility types were substituted.

RBF matched-pair health facilities randomly sampled for the household survey in the target nine provinces were compared with the final list of the selected BPHS health facilities. These matched-pair health facilities were added to the 2015 BSC sample to ensure inclusion of all matched-pair health facilities.

Data Collection

Data were collected from health facilities through two mechanisms – deployment of surveyors from Kabul, and training of local school teachers in provinces. For the relatively secure provinces, surveyors who were recruited and trained by SRTRO in Kabul were deployed to provinces to collect data from the sampled facilities. In the potentially insecure provinces a community-based method, which was piloted in 2009/10 and used since, was employed. According to this method local school teachers were trained on a concise version of NHSPA data collection instruments. The trained school teachers collected data while other school teachers were trained to do supervision and monitoring of the data collection. Therefore, in 2015, despite potential security concerns, local school teachers who were based in the respective community collected data in the potentially insecure provinces. The community-based approach was conducted in eight provinces in 2009/10, including Badghis, where both methods were used to test for reliability of the new approach, and it was found that the community-based approach is reliable and can be employed in insecure provinces. Thus, Kandahar and Kunduz, two insecure provinces which were excluded from the household survey are included in the health facility assessment.

Comparison with 2009/10 BSC

The BSC indicators and domains were revised in the 2011/12 round. The original BSC in 2004 was comprised of six domains and 29 indicators, in addition to three summary indicators. The revised BSC also has six domains, but they were restructured, and each domain has indicators which have either undergone revision or are completely new. For 2015, a new indicator concerning the status of health posts is included in the BSC. We use the 2009/10 BSC to serve as the baseline for this evaluation. Although the 2011/12 BSC is more similar to the 2015 BSC (used as the follow-up) in terms of questions and indicators, the 2011/12 BSC results would not provide an accurate representation of the baseline, or the pre-intervention performance as RBF was introduced to health facilities by 2011. Therefore, some of the 2009/10 BSC indicators that were no longer relevant or aligned with the 2015 BSC round were dropped and indicated in the tables. All indicators from 2009/10 were reanalyzed using the 2015 approach resulting in some being unmeasurable due to the lack of data at baseline.

Analysis

Primary analysis was undertaken as intention to treat at the health facility level. Two analytical approaches were conducted. The first analysis presents results on the same nine province matched-pair health facilities randomly selected as part of the RBF household survey. The second analysis presents the results of all 11 RBF province health facility performance. Exploratory data analysis was conducted prior to the main analysis. Items were checked for unusual values. Consistency of responses was examined when possible by comparing two or more variables that measure the same

characteristics such as age. Missingness was explored by study group to ensure it was not associated with treatment assignment.

For the nine provinces included in the household survey, data were analyzed using the same household survey weights and survey sample design for the health facility strata. A total of 61 matched-pair health facilities (122 total health facilities) were included at endline. The intervention effect was estimated using difference in means and 95% confidence intervals were calculated. Means were tested for both study groups using an adjusted Wald test.

For the total 11 RBF provinces, data were analyzed from all relevant health facilities, including health facilities from Kunduz and Kandahar, which were dropped in the household survey. This analysis provides a greater sample size, especially at baseline, and is illustrative of the performance of all RBF health facilities included in the BSC. Matched-pair analysis was not conducted for the 11 province health facilities since they were not randomly selected as matched-pairs. A total of 138 health facilities were included at baseline and 175 health facilities at endline. We then estimated means similarly for each indicator for each study group at baseline and endline and calculated 95% confidence intervals. Means were again tested for equality for both study groups using an adjusted Wald test. Since the health facilities were not randomly selected as part of a randomized comparison group trial as in the nine province analysis, we also conducted a difference in difference analysis. Three indicators were dropped from the analysis due to revisions in the scorecard after 2011/12 and these data were not collected at baseline. We used a difference-in-difference (DiD) approach for estimating the effect of the intervention while accounting for unrelated time trends. The analysis compares treatment and comparison groups at baseline and endline. We produced a DiD estimate and then tested the estimate as equal to zero using an adjusted Wald test. We present the DiD estimates for each indicator and overall mean score and its test p-value. A negative estimate is in favor of the comparison group and a positive estimate is in favor of the intervention group.

Table 1 shows the total number of health facilities sampled for both analytical approaches.

All data analyses were performed using Stata version 12.0 (13).

Table 28 Description of sample for analysis

	Baseline		Endline	
	Matched-pairs	Total	Matched-pairs	Total
9 provinces	22	44	61	122
11 provinces	-	138	-	175

Results

9 province analysis

Overall, the intervention group has a statistically significant higher score compared to the comparison group at endline, but not in absolute terms. The overall mean score for the intervention group is 56.7% and for the comparison group is 54.9%. The intervention group performed better than the comparison group on 14 indicators, of which seven indicators are statistically significantly different. These seven indicators are statistically significantly different at endline with higher performance by the intervention group: engagement of community in decision-making, health worker satisfaction, staff received training, equipment functionality, health facility management functionality, pharmaceuticals and vaccines availability, and universal precautions. We expect the RBF intervention to impact these indicators more directly compared to some of the other health facility performance indicators. Surprisingly, health worker motivation and overall client satisfaction showed no effect.

Table 29 Nine provinces – Matched-pair facility analysis from household survey sample

		2015 (n=122)		
		Intervention	Comparison	
Domain A: Patients and Community		mean (95% CI)	mean (95% CI)	p-value
1	Overall Client Satisfaction and Perceived Quality of Care Index	80.6 (78.6, 82.5)	79.4 (77.2, 81.6)	0.217
2	Community Involvement and Decision Making Index	91.8 (88.7, 94.9)	86.6 (81.6, 91.6)	<0.05
Domain B: Staff				
3	Health Worker Satisfaction Index	67.0 (65.4, 68.5)	64.9 (63.4, 66.5)	<0.05
4	Health Worker Motivation Index	72.5 (70.8, 74.2)	71.7 (70.2, 73.2)	0.507
5	Salary Payment Current	60.0 (51.3, 68.8)	56.7 (47.3, 66.2)	0.463
6	Staffing Index	26.0 (18.9, 33.1)	26.3 (17.9, 34.7)	0.953
7	Provider Knowledge Score	62.2 (60.6, 63.9)	62.1 (60.5, 63.7)	0.882
8	Revised Staff Received Training (in last year)	16.4 (13.9, 19.0)	13.2 (10.7, 15.6)	<0.05
Domain C: Physical Capacity				
9	Equipment Functionality Index	86.0 (83.8, 88.1)	79.5 (76.2, 82.9)	<0.01
10	Pharmaceuticals and Vaccines Availability Index	71.9 (67.7, 76.0)	68.5 (63.5, 73.5)	<0.10
11	Laboratory Functionality Index	80.7 (76.1, 85.3)	81.7 (76.7, 86.7)	0.673
12	Clinical Guidelines Index	85.1 (80.4, 89.8)	86.4 (83.0, 89.8)	0.595
13	Infrastructure Index	57.8 (52.5, 63.0)	54.1 (48.8, 59.5)	0.294
Domain D: Service Provision				
14	Client Background and Physical Assessment Index	81.5 (78.3, 84.6)	82.9 (79.9, 86.0)	0.282
15	Client Counselling Index	50.3 (45.1, 55.4)	48.5 (43.5, 53.5)	0.470
16	Universal Precautions	69.8 (66.2, 73.4)	63.3 (59.1, 67.5)	<0.01
17	Time Spent with Client	14.9 (9.8, 20.0)	15.6 (9.3, 21.9)	0.808
Domain E: Management Systems				
18	HMIS Use Index	79.2 (74.1, 84.4)	78.1 (72.8, 83.4)	0.704
19	Financial Systems	3.7 (0.29, 7.2)	1.8 (0, 4.0)	0.363
20	Health Facility Management Functionality Index	56.1 (53.1, 59.0)	51.1 (47.3, 54.9)	<0.05
Overall Mean Score		56.7 (55.4, 58.0)	54.9 (53.5, 56.3)	<0.01

11 province analysis

This analysis included all health facilities across the 11 provinces that introduced RBF. There are differences at baseline for 14 indicators, which are statistically significantly different. But we do not necessarily expect the two groups to be the same at baseline since they were not randomly sampled for this analysis, as they were in the nine province analysis above as part of the RBF household survey. The sampling for health facilities and households was done simultaneously, as households were selected within the catchment areas of the randomly selected matched-pair health facilities (please refer to the sampling design in Chapter 1 for more details. Due to insecurity, two provinces were dropped from the household survey data collection at baseline. Therefore, the nine-province analysis above includes the sampling design parallel to the household survey. This additional 11 province analysis includes all provinces that were included in the RBF program, including the two that were dropped from the baseline assessment. We therefore test the difference in difference estimates of the intervention and comparison groups in the total 11 provinces. At baseline, overall mean scores were also statistically significantly different between the two groups.

Overall, the intervention group has a statistically significant higher score compared to the comparison group in this secondary analysis at endline, as well. The intervention group mean score is 59.7% and the comparison group is 52.5% at endline ($p<0.05$). Both groups show overall improvements in scores, particularly the intervention group. The difference in difference estimates show that 13 indicators are positively in favor of the intervention group performance, whereas four indicators are in favor of the comparison group. Time spent with client is the only indicator with higher comparison group performance in the DiD that is statistically significant. The intervention group performs statistically significantly higher on salary payment current and infrastructure index. The greatest increase for the intervention group compared to the comparison group is on salary payment current with a difference of 16.7% ($p<0.05$).

Among the 13 indicators that the intervention group scored higher, the intervention group either scored higher or the same compared to its baseline score. The comparison group overall score declined since baseline and only nine indicators out of the 19 improved. The intervention group overall score improved since baseline.

Table 30 Eleven provinces – Unmatched-pair facility analysis from total BSC sample

		Baseline 2009/10 (n=138)			Endline 2015 (n=175)			DiD	
		Intervention	Comparison		Intervention	Comparison			
		mean (95% CI)	mean (95% CI)	p-value	mean (95% CI)	mean (95% CI)	p-value	Estimate	p-value
Domain A: Patients and Community									
1	Overall Client Satisfaction and Perceived Quality of Care Index	79.5 (77.1, 81.8)	77.2 (75.3, 79.1)	<0.01	81.0 (79.0, 82.9)	77.1 (75.0, 79.1)	0.158	1.6	0.463
2	Community Involvement and Decision Making Index	86.5 (80.3, 92.7)	78.3 (72.2, 84.4)	<0.05	91.7 (87.9, 95.5)	85.8 (82.1, 89.5)	<0.05	-2.3	0.624
Domain B: Staff									
3	Health Worker Satisfaction Index	71.3 (69.2, 73.4)	67.2 (65.3, 69.0)	<0.01	74.2 (72.7, 75.7)	68.7 (67.2, 70.2)	<0.01	1.4	0.474
4	Health Worker Motivation Index	-	-	-	-	-	-	-	-
5	Salary Payment Current	62.8 (54.4, 71.3)	63.5 (54.5, 72.5)	<0.01	65.4 (57.9, 72.8)	49.4 (39.7, 59.0)	0.913	16.7	<0.05
6	Staffing Index	29.3 (20.3, 38.3)	41.5 (30.4, 52.7)	0.189	32.3 (23.7, 40.8)	40.5 (30.1, 51.0)	<0.10	4.0	0.670
7	Provider Knowledge Score	-	-	-	-	-	-	-	-
8	Revised Staff Received Training (in last year)	12.0 (9.7, 14.4)	9.3 (7.0, 11.7)	<0.05	15.2 (12.9, 17.4)	11.4 (9.0, 13.7)	0.123	1.1	0.663
Domain C: Physical Capacity									
9	Equipment Functionality Index	78.3 (74.6, 82.0)	72.2 (67.6, 76.9)	<0.01	84.7 (82.4, 87.0)	75.3 (71.6, 79.1)	<0.05	3.3	0.300
10	Pharmaceuticals and Vaccines Availability Index	73.8 (69.0, 78.7)	79.5 (74.5, 84.4)	0.315	75.6 (70.9, 80.2)	78.4 (74.6, 82.1)	0.057	2.9	0.481
11	Laboratory Functionality Index	91.6 (88.2, 95.2)	58.5 (45.9, 71.0)	<0.01	86.2 (84.1, 88.4)	60.5 (48.7, 72.4)	<0.01	-7.4	0.410
12	Clinical Guidelines Index	82.1 (77.1, 87.1)	72.3 (65.6, 79.1)	<0.01	84.2 (79.9, 88.5)	70.1 (63.6, 76.7)	<0.05	4.3	0.409
13	Infrastructure Index	55.8 (49.9, 61.6)	53.6 (46.9, 60.3)	<0.01	62.8 (57.4, 68.2)	50.8 (44.7, 56.9)	0.622	9.8	<0.10
Domain D: Service Provision									
14	Client Background and Physical Assessment Index	77.9 (75.1, 80.7)	68.1 (65.5, 70.6)	<0.01	78.6 (76.1, 81.1)	66.9 (64.5, 69.3)	<0.01	1.9	0.474
15	Client Counselling Index	52.2 (47.8, 56.6)	32.4 (29.4, 35.4)	<0.01	52.4 (48.1, 56.7)	29.4 (26.6, 32.2)	<0.01	3.2	0.390
16	Universal Precautions	61.5 (57.5, 65.6)	73.3 (66.7, 80.0)	<0.10	66.2 (62.1, 70.3)	72.4 (66.8, 78.1)	<0.01	5.6	0.282
17	Time Spent with Client	21.3 (14.7, 27.8)	10.3 (6.5, 14.2)	0.638	15.9 (10.9, 20.9)	14.5 (10.8, 18.2)	<0.01	-9.6	<0.10

		Baseline 2009/10			Endline 2015			DiD	
		Intervention mean (95% CI)	Comparison mean (95% CI)	p-value	Intervention mean (95% CI)	Comparison mean (95% CI)	p-value	Estimate	p-value
Domain E: Management Systems									
18	HMIS Use Index	74.4 (68.8, 80.0)	71.9 (65.4, 78.4)	0.533	77.2 (72.1, 82.4)	75.4 (71.7, 79.1)	0.559	-0.7	0.893
19	Financial Systems	-	-	-	-	-	-	-	-
20	Health Facility Management Functionality Index	50.2 (46.5, 53.9)	49 (44.3, 53.8)	<0.05	54.0 (50.6, 57.4)	48.3 (44.8, 51.8)	0.703	4.5	0.242
Overall Mean Score		57.2 (55.5, 58.9)	53.4 (50.7, 56.1)	<0.01	59.7 (58.3, 61.1)	52.5 (51.1, 54.0)	<0.05	3.4	<0.10

CHAPTER 4: QUALITATIVE ASSESSMENT

Chapter 4 presents the results of the qualitative assessment of health facility staff and communities. Section 4.1 presents the health facility qualitative assessment and Section 4.2 presents the community level qualitative assessment.

SECTION 4.1: HEALTH FACILITY LEVEL

Methodology

The qualitative assessment was conducted in three provinces: Badakhshan, Bamyan, and Parwan. Two teams of five persons each were trained for a week on interview and group facilitation techniques. During the training, mock recorded interviews were conducted and interview guides were pilot tested. A total of 39 in-depth interviews (IDI) and focus group discussion (FGD) were conducted with health facility staff, including heads of facility, administrative staff, and clinical health workers across all BPHS level health facility types. A total of 8 comparison group health facilities and 13 intervention health facilities were visited across the three provinces. Health facilities were sampled through convenience sampling, by study group, performance levels (high and low performers), level of security, and within district representation to understand referral practices.

Interviewers obtained verbal informed consent from all participants. All interviews and focus group discussions were conducted in local language (Dari), voice recorded using a digital recorder, and later transcribed and translated into English. Transcripts were analyzed using content analysis, a qualitative data analysis method. All transcripts were initially reviewed for overall understanding and an integrated approach to coding was developed (17). Based on the initial review and the interview guide themes, we deductively organized a structure of initial code groups for analysis. Using this initial structure of codes, documents were analyzed line by line and additional emerging concepts were given new codes. Quotes from health facility staff interviews are presented by study group type.

Results

Community interaction

Coordination with communities through health councils and the presence of health posts and community health workers was discussed as an important channel for building patient trust and use of health facilities across both study groups. Health workers and staff reported that the low supply of medicines was an often discussed issue with communities as patients' complaints and poor satisfaction was due to the inability to access medicines at the point of service and rather being referred to the bazaar to fill prescriptions. Several health workers from treatment group facilities reported feeling that

their communities had high expectations that they could not meet due to the level of their health facility but efforts to communicate has helped to manage these expectations:

> *"Those who know the capacity of the clinic are satisfied, but those who do not have information about the capacity of a clinic have more expectations like a hospital… Nowadays due to giving information from the clinic and through CHWs and the health council about the clinic activity and capacity, people's knowledge changed and they understand that the capacity of a CHC clinic is less than a hospital." – Intervention*

Health system referrals and increased work load

A strong network of CHWs and health posts in communities ensured referrals to the BPHS clinics – this was true for both study groups. However, treatment health facilities reported that the provision of a 50 to 250 Afs. incentive for CHWs increased their motivation to refer a higher number of patients, especially pregnant women and TB patients.

> *"When there was not RBF project, the CHW did not try much as was saying that it does not matter if the women give birth at home or at a clinic. To be honest, in that time, we didn't try our best to emphasize bringing the pregnant women to the clinic for giving birth. But now because there is money, and this is also for the benefit of the people as well, CHWs try a lot. The women are told that giving birth is very difficult and it is risky and there is no need to wait at home. CHWs give them consultation to refer to a clinic whether it is at night or day time. We have night duty and an ambulance. This the main reason that referrals increased." – Intervention*

At the same time, similar perceptions of increased referrals by CHWs and the value of CHWs were also made by the comparison group health facilities.

> *"CHWs start their work from the villages where they do case findings. They refer the patients to the clinic…we provide services for the referred patients if we could give them treatment in our clinic and if not, we will refer them to other clinics or hospitals. We give them feedback about the services we provided. We tell them that the patient you sent received services for delivery, we helped them in immunization department and common illnesses in outpatient ward. We ensure them their patients were treated. We hold meetings once a month with the health posts and discuss our findings with them." – Comparison*

All health facility staff reported increased referrals of patients in terms of receipt of patients from lower level health facilities and referring out of patients to higher level health facilities. While all health workers interviewed reported having a general awareness of which of their neighboring health facilities had the RBF program or not, this was not a concern when referring patients.

> *"We do not care about the RBF program is implemented there or not, we will refer the patient to Takhar hospitals, Kunduz Hospitals, or Badakhshan hospitals." – Intervention*

> *"We have health posts around us that refer patients to us. Sometimes we refer the complicated cases to other health facilities. We refer the complicated and suspected cases of TB whose sputum was checked three times for taking xray or children with malnutrition. We know those health facilities and hospitals is under the coverage of RBF." – Comparison*

Health worker motivation

Staff were asked about their motivation to work and deliver services at their health facility. Motivating factors commonly mentioned across both study groups include training opportunities and increase personnel knowledge, appreciation acknowledgements from supervisors and the Provincial Public Health Department (PPHD), and overall satisfaction from patients on services received. The main sources of discouragement included salaries not paid on time, not having enough medicines available at the health facility, not having functional equipment, insecurity, and lack of transportation especially for immunization outreach programs. All RBF intervention health facilities, mentioned the RBF payments as a key motivating factor and payments in general as opposed to comparison group health facilities which had a greater focus on intrinsic motivations. Yet acknowledgment for performance and supportive supervision were important factors in motivating staff regardless of financial payments.

> *"Our duty and job motivates us to work well. Working is one of our community responsibilities and we signed the contract and are committed to our people and our country to continue our work." – Comparison*

> *"Regardless of the RBF project, ignoring good performance of a worker can lead to discouragement. For example, we eliminated a case of [specific disease] and we reached the area quickly without the need of other support. We quickly went to the area and prevented it from spreading. We should be encouraged from the projects, MoPH, etc…they should appreciate our work." – Intervention*

> *"Timely payment of salaries and RBF money and appreciation from the office encourages personnel. For example, the office should give appreciation letters to the personnel and encourage them for their good work." – RBF*

> *"Lack of timely payment of salaries, for example, it has been four months and we haven't received our salaries, which discourages the personnel." – Intervention*

> *"Everybody wants to cross the goals with are appointed. We are motivated to reach the goals of the PPHD." – Comparison*

Specific motivations referencing the RBF program are provided in the quotes below. Many RBF health facility staff reported increased motivation to perform their night duties and correctly document services as their payments depended on reaching targets and proof of performance.

> *"In the past when a woman gave birth at home, I did not write it in the book because there was no benefit to write. The money they gave us was enough for me but when I was informed on the RBF process, I registered in the book and received two thousand Afs." – Intervention*

> *"Without RBF we have worked in the past, but RBF money makes us more encouraged to work a lot and provide good services to the people. I can tell you most hospitals don't have RBF money, but they work for the people and for Allah." – Intervention*

"When RBF was not implemented here in this clinic, we were not interested to come to the clinic during the night for delivery. After RBF, we stayed in the clinic all night because RBF program pays us, not only in MCH in other parts we also can see the efficiency of the RBF program." – Intervention

General RBF knowledge and perceptions

Comparison group health facility staff were asked about their understanding of the RBF program and selection for participation, as well as whether knowledge of the RBF program in some health facilities but not in their own affected their daily work. All comparison group health facility staff were aware of the RBF program and the program specific targets, but did not have a clear understanding of the randomized selection process. All treatment health facility staff showed higher knowledge of RBF details such as the randomized selection process, objectives, targets, and payment processes. Staff who had worked since the introduction of RBF recalled being oriented to the program and its objectives. New staff had less knowledge about the program, however, and were unaware of targets. All health facility staff from both study groups were well-aware of the verification processes.

All health facility staff expressed high expectations of an expansion of RBF – either expanding to other service areas in the current intervention health facilities, or expanding to also cover comparison group health facilities. Although the intervention health facility staff expect to see an expansion of RBF to other service areas, they also expect an increase in the payment amounts.

"Our expectation from the RBF program is that they send the incentive on time, and increase our incentives." – Intervention

"Our expectations from the RBF program is mainly to continue. People need this encouragement. They are doing night duty and referring obstetric patients at night now." – Intervention

"They should continue implementation of this program. We want these other programs also to be included in the RBF program, like OPD, laboratory technician, pharmacy, and nutrition." – Intervention

RBF health facility staff perceived the program to be positive and effective in increasing the quality the quantity of services as well as in encouraging staff. There was some reported negative consequences such as disputes between doctors and midwives on differences in the payment disbursement amounts, but that these were resolved over time. Some health workers reported being motivated by competing with other RBF health facilities by comparing RBF payments received and with comparison group health facilities in performance for fear of losing the RBF program.

"I think the [comparison group health facilities] will be more interested to come under the RBF program, so they want to work very hard and we must be careful of our work to not lose the program which means we are both motivated and competitive." – Intervention

"We have positive competition with [other RBF health facilities]. One clinic received around three hundred thousand Afghanis more than us and we regretted why we did not gain the same amount like them. If only we worked more than usual to receive more money from RBF is the general impression so it encourages us to work hard and gain a lot more." – Intervention

Targets, competition, and expectations among comparison group health facilities

Comparison group health facility staff were specifically asked how they felt about RBF being implemented in other health facilities and not their own, particularly as they showed poor knowledge about the selection process. Staff expressed misunderstanding for why they were not included in the program and future expectations of participating in the program and receiving additional payments if they performed well and reached their targets. Some health facilities expressed explicit expectations of receiving the program in the future, while others valued the importance of reaching their targets regardless of whether they had the RBF program. None of the staff explicitly mentioned competing against their RBF implementing peer health facilities, but rather, they perceived the need to prove themselves and their ability to reach targets in the hopes of also participating in the program.

> *"I asked them [NGO] why we are not included in RBF and if it is that our functionality is poor. He said no, our functionality is good, but only a couple of health facilities were selected as a sample so only these sample will receive RBF if they reach targets. Our health facility is not included in the program but in the coming years, if we reach our targets then we would get an amount of money for it, but we are not included in the program." – Comparison*

> *"As my colleagues mentioned, the RBF program will take a facility under coverage if we cross the targets set by the MoPH in PENTA3 and deliveries so our clinic will come under treatment of RBF. Many times we have cross the targets but we still don't know when our clinic will come under the coverage of the RBF program. They told us that if you cross the targets, your clinic will come under the RBF program." –Comparison*

> *"I have no special opinion [about not having the RBF program]. We are not unhappy, but we try to join the RBF program and to reach the goals in order to come under the coverage of RBF." – Comparison*

> *"Whether our clinic comes under the coverage of RBF or not, we have to reach the goal of the PPHD because we have a responsibility against the salary we are taking from the government. The money of RBF is not very important for us, but if RBF program is implemented in this clinic that is very good." – Comparison*

Quality and quantity of services

Positive changes in the quality and quantity of health services over the past five years were reported by both RBF health facilities and comparison group health facilities. Specific areas of improved quality include maternal care, vaccinations, and TB detection and treatment. Both study group facilities mentioned reaching targets for these service areas as justification for their perceived improvement in quality. Some health facility staff also recalled different projects that were introduced such as the URC quality improvement project and specific projects related to maternal health and nutrition.

> *"You know there aren't enough medicine and facilities in the OPD section. The people are complete satisfied from two sections, though, delivery section and vaccination section because we provide services to them on time." – Intervention*

"This project has brought good changes in the delivery section. The health workers in the MCH office such as midwives are trying to work better and provide better services to patients and this shows with higher data at the end of the month. We have more services provided in the MCH office, like deliveries, antenatal care, and prenatal care. Detection and diagnosis of TB cases is also in the RBF project which has encouraged the laboratory technician to work more than his target set by the clinic and get more incentives according to his performance." – Intervention

"By the name of God, the changes that have come in the clinic are a lot and both in quality and quantity. The health facility has staff and clients which have increased compared to the past in obstetrics and vaccinations. In OPD, day by day, we are witnessing changes in the graph of targets which is rising. Most of the deliveries are taking place in the clinic, no child is left without vaccination whether the delivery is done here or not in the clinic. Detecting TB has improved and the patients are followed-up and proper treatment is given to them. These are the changes we see." – Comparison

Challenges to delivering high quality services were also reported by both study group health facilities. The main challenges included the lack of female doctors, overburdened midwives, lack of drug supplies, lack of transportation for outreach, and poor patient awareness of appropriate drug use.

"There is a lack of female doctors in this clinic. It has been six months I have been working in this clinic yet there isn't any female doctor and I am performing both duties (midwife and doctor). The midwives observe patients for ANC, PNC, and deliveries. There is also a lack of medicine. We have a supply of drugs every three months. The percent of anti-biotics is not high. Our community is a backward society and our kids have diseases and other things and no matter what they try to use more anti-biotics. We try to lower it and give less to our patients but it is a major problem in regards to drugs and satisfaction of the patients." – Intervention

"People have very low health awareness and they can't observe their personal hygiene which is why they are always sick. The usage of anti-biotics is high but the usage of anti-biotics in the BPHS project is low and they focus more on people's awareness. We respect this principle but people have lower health awareness and need drugs until their awareness level is improved. Those people with higher level of awareness and higher literacy observe their personal hygiene more. Anti-biotics are not good for patients with first stage of diarrhea but people want us to prescribe them and whenever we do not give them their desired medicine, they refer to private clinics which has an economic impact and the private clinic's work becomes greater compared to our clinic." – Intervention

"Praise be to Allah because the quality of services are quite good but unfortunately the main problem is the lack of medicines here. I have no received any medicines this year [1394]. People are referred to the clinic for medicine. When emergency patients come, I prescribe out of the clinic because there are no medicines." – Comparison

"We do not have transportation for outreach. For example, when traveling three hours, vaccine carrier is with you with nine types of vaccines and reaching the village on time is difficult because two types of vaccines have only six hours and after that it is useless. There is great waste in the vaccine section." –Comparison

RBF health staff mentioned that while the quality of health services are perceived to have improved, stopping the RBF program would be result in unmotivated staff, which may affect the existing quality of care.

> *"The quality has changed and gotten better but if the RBF program stops the incentives then we are not able to encourage the CHWs so they will be unmotivated and not work properly." – Intervention*

RBF accountability mechanisms

Many health facilities under RBF reported receiving payments according to their documented performance, with consistent and transparent processes. Many health facilities implementing RBF reported having a representative committee that made decisions for fund use and disbursements. However, there were still common reports of poor transparency, and inconsistent practices, particularly with the changes of NGO implementers. Some health facilities reported a hierarchical process, rather than a representative committee, for decision making, as well as less facility-based autonomy due to decisions made directly by the implementing NGO. The RBF intervention relies on consistent payments that are timely and sufficient in the agreed amounts for health workers. The logic of the intervention rests on a change in behavior in response to the incentive and a change in the system of service delivery, but this would not be possible if the payments were not received on time or in the amount promised.

> *"If we consider RBF program, the most important things for motivation is increasing the RBF incentive, supply of drugs on time and also the equipment which are out of order. They should supply or change the shortage of items so that we can offer good services for the people. Another important component is the transparency of the RBF program. One midwife can understand how to achieve the target. If she had 30 deliveries she will know 15 deliveries of the target was completed and the incentive of 15 other deliveries will be paid for her from the RBF program, but the problem now is there is no transparency in the RBF budget." – Intervention*

> *"Before, another organization the implementer and during that time the RBF incentive was received very late but was in accordance with the reports. Now, with [new] implementer, we did not receive a quarter of our achievements, meaning they send us less." – Intervention*

> *"Before there was good transparency in RBF finance and we were able to understand the percentage of our reports compared with the money we received, but now we do not know how much the target is, where they get it, and the money they send does not match with our reports." – Intervention*

Fund decisions and use

Most of the RBF health facility staff reported having a committee with different staff representatives to decide how the RBF funds should be disbursed and used. Some health facilities reported a very hierarchical approach to decision making by the implementing NGOs. Many health facilities reported the importance of including CHWs as recipients of funds in order to increase referrals to clinics. Many health workers thought midwives were the most important position for the program in order to reach

their targets but that their incentive amount was too low. In terms of payment receipt, many health facility staff reported frequently late payments of their salaries as well as their RBF incentives, experiencing up to 4 to 9 months in payment delays.

RBF verification

The RBF verification process of quarterly monitoring and follow-up of HMIS reports from the registries to the households was repeatedly mentioned by both study group staff. Verification brought about issues of concern for how health facilities are penalized as well as benefit and value. Both intervention and comparison group health facilities discussed how the routine verification changed health workers' behaviors to be more attentive to documentation and reporting of patients.

> *"His supervision makes us pay attention to be punctual in our work and to be very careful. For example our registration is the same as our reports, and it means the report would not be wrong." –Comparison*

> *"One of the issues which was the means of encouragement, evaluator team had come after three months but now they come after six months which is also disappointing" –Intervention*

Yet, treatment health facilities mentioned challenges to the verification process, particularly when monitors are unable to identify the patients in communities. Health workers felt that if a patient does not provide accurate information regarding their name or contact information, then they should not be considered incorrect reporting, as they get penalized.

> *"RBF money requires the names of the patients to be the exact names registered at birth. For example, this month there were eight births and different health services were provided for the eight women, but some of these women didn't provide their exact names. They provided one name to the clinic while her real name was something else in the village which created misunderstanding. As a result, this caused a mismatch and only four or three women were counted and paid. The [monitors] also claimed that these four were the real and four others we reported were wrong. It is not wrong. This is the fault of people when they provided wrong names and addresses as they said the wife Mullah in this area where there are lots of Mullahs, so which Mullah?" – Intervention*

> *"Another negative point is for example, the pregnant woman when are referred here to the clinic they cannot tell the exact name of the area where they have come from when a midwife asks her where she is from. Say I have come from one village, while there is another village with a similar name. When [monitors] visit they ask there is another by the same name when they are asking about someone there they cannot find the wanted person, while this person have come from another village. It happened in the last team visit, I was with the doctor and there was a patient who had come from one village but gave the name of the village as something else. When they visited they couldn't find the wanted patient but when they went to another village, they found the patient there because the patients who come from the villages are uneducated. It is also another element which causes discouragement of the staff here." – Intervention*

The verification process seemed to have a positive effect on comparison group health facilities, in particular. Comparison group health facility staff reported improved documentation and efforts to fix data entry errors, as well as becoming more aware of their service delivery targets. They also perceived

an added value to monitors' visits as households may have seen their presence in communities as a validation of the clinic's services and increasing their reputation and trust.

> *"The evaluation of the third parity supervisor had a direct and helpful effect on our work. Their evaluation alerted us to be punctual in registration and reports. Their monitoring and evaluation made our errors appear in registrations and also we could evaluate ourselves to see if we reached the goal or not." – Comparison*

> *"Their monitoring and evaluation had a direct effect on our work quality also when they take out the name of some patients for vaccine or delivery in this clinic and they go to the villages. They first ask to find the people. Their methods and this program enhances trust by people about the functions of the health center, after that the peoples rush for taking vaccinations and there is a rush of pregnant women for ANC, Delivery and PNC. All will be increased." –Comparison*

> *"The monitors and evaluators come here and they all talk about RBF or non-RBF program or vaccination or delivery or pharmacy. We will maximize the usage of their monitoring and evaluation because it has positive direct effect on our work for example before there was one or two delivery taking place and when we receive the number of targets we become more active to reach the goals set by the MOPH." –Comparison*

SECTION 4.2: COMMUNITY LEVEL

Methods

The 2015 community level qualitative assessment aims to provide community level perceptions of changes in the quality of health services. Specifically:

- Perceived changes that occurred in the last five years, since the RBF program was introduced in health facilities. We asked about the changes that occurred in communities and the reasons for the change.
- Perceived quality of health care services, challenges, and opportunities. We asked for participants' views on the challenges and opportunities with health services over the past five years.

Similar data collectors conducted focus group discussions with community groups in the catchment areas of the selected health facilities in the qualitative assessment. Three target groups were formed in communities – community stakeholders or leaders, men's group, and women's group. The community stakeholder group was formed by a community leader and other members of position such as CHWs or teachers. A male and female couple was selected randomly from households to participate in the gender-specific groups. Participants were randomly selected in the communities using a transect walk, selecting a random number, and inviting every nth household couple to participate. One village from the catchment area of five comparison group health facilities and five intervention group health facilities in Badakhshan, Bamyan, and Parwan were included in this qualitative assessment. Focus

group discussions were held with a total of seven community stakeholder groups, 10 men's groups, and 10 women's groups (27 discussions total).

The focus group discussions with communities were designed based on a group-approach to data collection, based on earlier developed PADev methodology (see for more details and the guidelines for this method www.PADev.nl). This is a structured method of organizing focus group discussions, aimed to limit and pre-structure data collected through discussions with community members.

Results

Positive and negative changes in quality of health services

The perceived positive changes most often mentioned in treatment villages were the vaccination programs (9), family planning services (5), the quality of the deliveries services (6) and the fact that the clinic had been built in the first place in the past few years (3). In comparison group villages, these same points were mentioned: vaccination (10), OPD (5), deliveries (3) and the building of the clinic itself (3). The availability of drugs and the quality of emergency patient care were also mentioned in treatment villages (both 3 times). In comparison group villages, out-door patient care was mentioned as having been improved. Table 31 shows the frequencies of positive changes reported by study group.

Table 31 Summary table of positive changes by study group

	Vaccination	Deliveries	OPD	Family planning	Clinic established	ANC/ PNC	Medicines	Emergency care	Other	Total
Intervention	9	6	3	5	3	2	3	3	6	40
Comparison	10	4	7	2	4	2	0	0	5	34
Total	19	10	10	7	7	4	3	3	11	74

Note: other = all issues with score of 1 in either treatment or comparison group villages.

In general, the quality of health care is perceived to have improved. Women's, men's, and stakeholder groups all considered vaccinations and deliveries as having improved the most, and having had the most impact. Women's groups commonly mentioned the availability of 24 hour services for delivery and availability of ambulances as positive changes.

Though far fewer in number, negative changes were also mentioned by both study groups. Table 32 provides a summary of the negative changes mentioned. Negative assessments were most severe towards the quality of the service, not towards the availability of staff or supplies.

Table 32 Summary table of negative changes by study group

	Medicines	Ambulance	**S**taff availability	Clinic established	**D**eliveries	**L**ab facilities	**S**taff attitude	**O**ther	**T**otal
Intervention	0	0	0	1	0	0	0	2	3
Comparison	5	4	4	2	2	2	2	6	27
Total	5	4	4	3	2	2	2	8	30

Challenges and opportunities

Community groups were asked about general challenges they faced, not necessarily specific to health. A main challenge commonly reported was the lack of roads and transportation options to travel and transport patients to and from health centers. Other community challenges reported were lack of electricity, weak economy, and insecurity. While rankings differed by group, these were most commonly mentioned by all. Health-related challenges were not commonly reported as the most significant. The main opportunity identified was education through schools to increase overall health awareness. Table 33 shows the rankings of challenges by province and study group.

Table 33 Summary table of rankings of challenges

Badakhshan	Intervention	Comparison	Average	Difference
Health care quality	3,7	2,2	3,1	1,5
Drinking water quality	6,0	4,0	5,0	2,0
Infrastructure quality	3,2	4,5	3,6	-1,3
Economic situation	5,1	4,8	4,9	0,3
Parwan				
Health care quality		3,7	3,7	
Drinking water quality		4,0	4,0	
Infrastructure quality		4,0	4,0	
Economic situation		6,0	6,0	
Bamyan				
Health care quality	3,5	3,8	3,6	-0,3
Drinking water quality	3,2	1,7	2,4	1,5
Infrastructure quality	3,4	5,2	4,4	-1,8
Economic situation	3,5	5,3	4,2	-1,8

Note: Economic situation includes security and school quality. Low scores indicate a highly problematic situation. The color is given on the basis of a scale of 9 (1-3 is red, 4-6 is yellow, 7-9 is green).

CHAPTER 5: DISCUSSION

Large improvements are seen across all health indicators since baseline. Both intervention and comparison groups have improved in coverage of key maternal and child health services in 2015. However, the results of the RBF household survey indicate that the intervention did not have a statistically significant positive effect overall. Where statistically significant differences were found, changes were greater in the comparison group, such as in the utilization of outpatient services for adults. Although the intervention group showed higher absolute proportions compared to the comparison group across all health indicators except for modern contraceptive use, these differences were not statistically significant. Furthermore, the comparison group tends to be more pro-poor at endline, although they were less pro-poor at baseline. The greatest changes are observed when we remove the experimental design and analyze the data as a pre and post intervention. The statistically significant improvements since baseline indicate that time is likely a significant predictor of change. As the intervention has been implemented over a period of five years, the impact of time is not surprising. When we comparison group for time in the difference-in-difference analyses, we find that the intervention does not have a positive effect.

Overall, the intervention and comparison groups were similar across village characteristics spanning economic activity, external shocks, women's rights, governance, and basic service availability. There were some indicators that showed statistically significant differences between the two study groups but most were unlikely to explain the RBF household survey results. The main difference that might have some effect is the difference in the availability of public and private hospitals for the intervention and comparison groups. The intervention group had greater access to public hospitals but the comparison group has greater access to private hospitals. Yet, the availability of both public and private hospitals is limited for both study groups so the chance of this having a considerable differential effect on the two study groups is also unlikely. Both communities and health facility staff interviewed mentioned poor roads and lack of transportation as barriers to accessing health services. Limited infrastructure in communities creates delays in care seeking or prevents access, especially when coupled with insecurity, for women and children.

It is unclear why no differences were found between the intervention and comparison groups at endline. One reason why no treatment effect was detected may be due to the issue of spillover resulting from the design of the overall intervention. The sampling approach was designed to randomly select health facility matched-pairs by health facility type. For example, basic health centers were matched with other BHCs and randomly selected as a pair within the BHC facility type strata. This approach did not take into consideration the interaction of health facilities across the health system from a SC to a BHC to a CHC and finally to a DH. Health facilities do not operate independently within the health system, rather they are designed to interact by increasing referrals upward as necessary and by design, operate within overlapping catchment areas. We investigated the issue of spillover further by reviewing the full list of health facilities in either treatment group in all nine provinces from the HMIS database. We observed that health facilities by type, district, and treatment assignment often fell within the same sub-health system. In a single district, a DH may be assigned as a comparison group health

facility while the CHCs are mixed comparison group and treatment as well as the BHCs and so on. Another component of spillover may include the implementing partner for service delivery, often responsible for both treatment and comparison group health facilities. Some evidence for possible spillover is presented in the qualitative assessment where health facilities had knowledge of treatment assignment of facilities around them but regardless, referred based on patient needs. The comparison group health facility staff also reported focused efforts on reaching MoPH targets in the same services areas as RBF objectives – maternal and child health as well as TB. Reaching targets in general, with or without the financial incentive, may have been encouraged for all facilities by the implementing NGO. It is not clear whether the targets were the similar for all health facilities, regardless of RBF assignment, but it would not be surprising since implementation of RBF and BPHS services were by the same NGO.

At the health facility level, the estimated difference in overall mean score between the intervention and comparison groups from baseline to endline was 3.4% ($p<0.10$) for the 11 province analysis and 1.8% difference between intervention and comparison group groups for the nine province analysis ($p<0.01$). While these differences are statistically significant, they are not very large and possibly not truly meaningful in understanding the performance differences of the two study groups. Indicators where differences were found show a more positive picture for the intervention.

In the nine province analysis, community involvement and decision making index (indicator 2) showed a 5.2% difference, health worker satisfaction index (indicator 3) showed a 2.1% difference, staff training (indicator 8) showed a 3.2% difference, equipment functionality index (indicator 9) showed a 6.5% difference, pharmaceuticals and vaccines availability index (indicator 10) showed a 3.4% difference, universal precautions (indicator 16) showed a 6.5% difference, and health facility management functionality index (indicator 20) showed a 5% difference. These differences on average account for a 5% higher score for the intervention group compared to the comparison group. The DiD estimates show fewer statistically significant differences between intervention and comparison groups between baseline and endline: salary payment current 16.7% (indicator 5), infrastructure index 9.8% (indicator 13) for the intervention group and time spent with client -9.6% (indicator 17) for the comparison group. These results indicate that the RBF intervention had an effect at the health facility level which was larger in specific aspects of performance and health facility functions rather than overall.

We did not find an effect, however, on health worker motivation and client satisfaction. Health worker motivation may be influenced by various factors besides financial incentives. Interviews with health workers showed that while the RBF program was considered a good additional motivator, salaries and the incentive were not always received on time. Though, late payments cannot be confirmed with our data, delayed payments or long periods between payments, particularly for RBF may defeat the original intent as they are not as directly linked with staff performance as they should have been. Additionally, the comparison group health facility staff reported being motivated to achieve their own targets for performance. And both groups reported many demotivating factors as well such as a challenging work environment.

The lack of difference in overall client satisfaction is not entirely surprising as the RBF household survey showed no differences at the household level between the two groups on health outcomes. However, there was a minor improvement in client satisfaction in the intervention group since baseline but it is not statistically significant. This is surprising since utilization of maternal and child health care services has increased, as seen in the household survey. Although, general utilization of outpatient services has not really increased. A possible explanation is the limited supply of drugs and medicines at the BPHS clinics, resulting in patients purchasing their prescriptions in the private sector, a challenge commonly reported by health workers in interviews.

The RBF intervention seems to have had an overall positive effect at the health facility level through the BSC as well as through health worker perceptions, yet these improvements have not translated to the population level. At the same time, both study group health facility staff perceived overall positive changes in the quality, quantity, and motivation of health workers, with or without participation in the RBF program.

Another possible source of bias may be due to the concept of competition among comparison group health workers. As a pragmatic trial, blinding was not possible. Health workers from either treatment group were aware of the RBF program and their treatment assignment. Health workers in comparison group health facilities may have been motivated to improve the quality of services with the expectation that they too would become eligible for the financial incentive. This misperception was shared during our interviews with health workers from the comparison group. They not only acknowledge the need to meet their own targets set by the NGO/MoPH, but also they acknowledged wanting to improve their performance for the change to participate in the RBF program. The comparison group health facility showed poor knowledge and understanding of the RBF program design compared to the health workers receiving the intervention. The MoPH and implementing NGOs should have better managed such expectations. Additional expectations of the continued program by the RBF health workers will need to be managed as well. Appropriate communication with health facility staff on the RBF program and any future changes should be well managed for both study groups to prevent demotivating staff as much as possible.

Verification of HMIS reports in both treatment groups may have also diluted the effect of the intervention as ongoing monitoring and quality checks of health facility reports may have motivated improvement of services in comparison group health facilities. Health workers from both study groups expressed improvements in reporting and documenting services and patients as required by the HMIS because of the verification process. Health workers from the comparison group also perceived positive consequences of verification such as the influence of monitoring visits on the clinic's reputation in communities. This positive influence was perceived to increase patients from visited communities. The intervention group also reported positive perceptions and use of the verification process. At the same time, challenges to the verification system were also reported. The change in verification from every three months to every six months was reported as a disappointment by intervention health workers. Additionally, the punitive results of the verification process were reported as demotivating when monitors are unable to identify the same patients in the community

and those patients are not counted towards the target – health workers defended themselves against supposed incorrect reporting in the HMIS by blaming the patients for not providing their correct names and contact information, a situation perceived to be out of their comparison group.

There may have also been a question of intervention fidelity – was the RBF intervention implemented as designed? Interviews with health workers provide some information that intervention fidelity was challenging during implementation. Were health workers in the intervention arm aware of the incentive program and their expected outputs in service delivery and quality? Health workers seemed to be aware of their targets but health workers in the comparison group seemed to also have targets, possibly similar targets due to the same implementing NGO management. Did health workers receive their expected payments, were the payments accurate, and were the payments administered in a timely manner? Payments were not made on time nor were they reported to be sufficient. Alternatively, there may be unobservable factors affecting the comparison group and not treatment group, resulting in improvements in the comparison group health facilities outside of the RBF intervention. We cannot be certain that any of these are the cause of the limited effect of the intervention, however, they do invalidate the theory of change expected from the RBF intervention as it was implemented.

Another possible design issue may be related to the matching approach and identification of control (comparison group) health facilities. Health facilities were matched by OPD visits the year before the introduction of the RBF intervention. Matching on this characteristic alone may not have been sufficient for identifying appropriate comparison groups for intervention health facilities.

The RBF provinces by endline have shown significant improvement on key maternal and child health indicators. It is possible that a combination of factors contributed to these improvements in both study groups, but communities were overall similar to one another in terms of development and practices. Improvements at the health facility level were not as dramatic but differences were found between the study groups. Setting clear targets for health facilities and monitoring achievement through the HMIS verification process seemed to have the greatest impact on both study groups. Sources of motivation and whether RBF payments influenced health worker motivations for better performance remains ambiguous – while perceptions were positive, actual measures showed no effect – additionally, delayed payments for salaries and RBF incentives may have diluted a more direct and positive effect.

While this study was designed as a randomized comparison group trial, which is considered a gold standard in assessing impact and attribution of an intervention, there are several limitations that should be considered. There is always potential for selection bias and response bias in a cluster sample. Additionally, due to insecurity, we were not able to visit 13 of the original clusters – statistical analysis based on observed characteristics showed that these clusters were not different at baseline, however, there may be bias due to unobservable characteristics unaccounted for in the data. Previous RBF programs such as the case of Rwanda, randomized the treatment for health facilities at the district level. We believe that the design of the RBF program in Afghanistan with treatment assignment at the health facility type level coupled with implementation and management of health facilities in both

study arms by the same NGO resulted in significant spillover effects. This, in addition to HMIS verification for both study groups resulted in a lack of a true control group to assess the treatment effect. We believe the lack of this true control group is the largest limitation of our study. A potential fix for this may be to generate a synthetic control group using data on the rest of the country.

The MoPH is currently implementing the System Enhancement for Health Action in Transition (SEHAT) project, for which RBF was intended to be a promising financing mechanism for linking NGOs to health facility performance. This pilot project was shown to be ineffective for the various reasons discussed above. Health facilities seemed responsive to the HMIS verification process, which may be an alternative method to assess performance, rather than payments.

CHAPTER 6: RECOMMENDATIONS

The MoPH, international partners, and implementing NGOs will need to consider how best to move forward in light of the impact evaluation findings. Based on the findings, our recommendations include the following:

- **Communicate clear targets to all health facilities implementing the BPHS and EPHS**: All health workers in the public sector should be made aware of the MoPH's goals, which are translated into feasible and attainable targets over time. Aiming for targets and achieving them can motivate health workers to feel as part of the team and as part of the greater system. The RBF program aimed to make targets and achievements very clear for health workers. While some health facilities reported confusion, health workers who clearly understand their performance targets, expressed certainty in what they were expected to deliver. This was true for both study groups. The RBF payments, of course, provided some additional motivation to health workers, but motivation through target setting and monitoring on the comparison group health facilities seemed to occur without additional payment.

- **Maintain and improve the HMIS verification**: Continued feedback through the verification process is important for health workers to understand their performance levels and where they need to specifically improve. This feedback mechanism is only possible through proper documentation. Monitoring should occur more frequently than not – semi-annual verification and feedback may not be as influential as quarterly verification. A mix of callbacks to patients using mobile numbers and field visits/in-person spot checks may help to balance the need for routine and rigorous monitoring with the financial cost and risk of insecurity of traveling. Health workers may need increased engagement in the HMIS verification process, to ensure that they feel the process is fair and valid, and that they are not being penalized or accused of lying. Mechanisms for ensuring that they collect correct names and contact information should be in place and communication of what they are responsible for in the documentation process is necessary.

- **Improve monitoring and supportive supervision**: While HMIS verification is an essential process in monitoring, greater supportive supervision and use of the MoPH monitoring checklists with appropriate feedback to health workers/health facilities is needed. A transparent process with the checklist data available and accessible in a database would help health workers to understand their progress. Monitoring checklists and feedback forms/communication strategies should be reviewed and updated periodically to prevent health workers from performing to the checklist.

- **Consider alternative mechanisms for motivating health workers**: Based on annual performance measures through the BSC, HMIS verification, and monitoring, the MoPH and PPHD can provide annual achievement certificates to health facilities based on agreed criteria by stakeholders. This is a simple and cost-effective method to motivate health workers.

- **Ensure timely payment of salaries**: Health facility staff in both study groups reported long delays in the disbursements of salaries. While they recognized that salaries are low, and they would prefer higher salaries, they also acknowledged their commitment to a position with a certain salary amount. Yet, long delays in salary disbursements are demotivating to an already challenging role. Ensuring timely payment of salaries is an overall government challenge – mobile payments or other innovative mechanisms may provide solutions.

- **Consider alternative demand side interventions**: The lack of available transportation for women and children to seek care at health facilities was mentioned by both communities and health facility staff as a barrier, yet, where ambulances were available, this form of transportation was perceived as contributing positively to the quality of care. While an ambulance for all clinics is not feasible or necessary, an innovative transportation intervention should be considered such as partnering with private drivers, providing vouchers, using mobile payments, or harnessing other local modes of transport.

- **Conduct additional qualitative studies to improve the verification process**: The qualitative assessment under the RBF impact evaluation aimed to obtain general perceptions and understanding of the RBF program. This experience has exposed additional areas of further investigation, particularly in verification, to improve and strengthen the process and its acceptability. For instance, understanding views of verification by mobile phone, collection of mobile phone numbers of clients, and frequency of verification are all areas that need to be better understood for improving the verification process.

- **Organize a stakeholder workshop to understand the RBF impact evaluation results in its entirety and to decide how to move forward**: While there is no strong evidence of the intervention's effectiveness against its main objectives on population health, there is some positive effect on health facility performance and a perceived benefit by health workers. Any decision made by policymakers on the continuation, expansion, or completion of the intervention should be communicated to all stakeholders considering the mixed evidence.

- **Prepare a de-implementation plan and communication strategy if RBF is not to be continued or expanded:** If RBF is not expected to continue or expand to other comparison group health facilities, a de-implementation plan and communication strategy is needed. Health workers have expressed certain expectations and hopes from the program. To prevent demotivation and disappointment among a large group of health facility staff – from guards to administrators, to clinicians, to CHWs – a clear plan for de-implementation and a communication strategy are needed.

REFERENCES

1. Afghanistan Public Health Institute, Ministry of Public Health (APHI/MoPH)[Afghanistan], Central Statistics Organization (CSO) [Afghanistan], ICF Macro, Indian Institute of Health Management Research, World Health Organization Regional Office for he Eastern Mediterranean. Afghanistan Mortality Survey 2010. 2011.
2. Afghanistan Health Survey (AHS) 2006: Estimates of Priority Health Indicators for Rural Afghanistan. Kabul; 2006.
3. Bartlett L a, Dalil S, Salama P, Mawji S, Whitehead S. Conceiving and dying in Afghanistan. Lancet [Internet]. 2005;365(9476):2006. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15950713
4. Honda A. 10 Best Resources on ... Pay for Performance in Low- and Middle-Income Countries. Health Policy Plan [Internet]. 2012 Sep 7 [cited 2013 Mar 12];(2009):2009–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22962058
5. Basinga P, Gertler PJ, Binagwaho A, Soucat ALB, Sturdy J, Vermeersch CMJ. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. Lancet [Internet]. Elsevier Ltd; 2011 Apr 23 [cited 2013 Feb 27];377(9775):1421–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21515164
6. Ashir GM, Doctor HV, Afenyadu GY. Performance Based Financing and Uptake of Maternal and Child Health Services in Yobe Sate, Northern Nigeria. Glob J Health Sci [Internet]. 2013 Jan 29 [cited 2013 Apr 14];5(3):34–41. Available from: http://www.ccsenet.org/journal/index.php/gjhs/article/view/23576
7. Looij F Van De, Doorslaer E Van. Introduction Of Performance- Based Financing In Burundi Was Associated With Improvements In Care And Quality. Health Aff. 2014;33(12):2179–87.
8. Toonen J, Canavan A, Vergeer P, Elovainio R. Learning Lessons on Implementing Performance Based Financing, from a Multi-Country Evaluation. Amsterdam; 2009.
9. Oxman AD, Fretheim A. Can paying for results help to achieve the Millennium Development Goals? Overview of the effectiveness of results-based financing. J Evid Based Med [Internet]. 2009 May [cited 2013 Apr 14];2(2):70–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21348993
10. Oxman A. An overview of research on the effects of results-based financing. 2008.
11. Ministry of Public Health [Afghanistan]. Results-based financing project in Afghanistan: Baseline household survey report. 2010.
12. Hayes R, Moulton L. Cluster Randomized Trials. Boca Raton: CRC Press; 2009.
13. StataCorp. Stata Statistical Software. College Station, TX: StataCorp LP; 2011.
14. Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Health Policy Plan [Internet]. 2006;21(6):459–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17030551
15. O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M. Analyzing Health Equity Using Household Survey Data. World Bank Publications; 2007 Oct 27; Available from:

http://www.worldbank.icebox.ingenta.com/content/wb/2498

16. Ministry of Public Health [Afghanistan], Silk Route Training and Research Organization (SRTRO), Royal Tropical Institute (KIT). The Balanced Scorecard Report: Basic Package of Health Services 2015. 2015.

17. Bradley EH, Curry L a., Devers KJ. Qualitative data analysis for health services research: Developing taxonomy, themes, and theory. Health Serv Res. 2007;42(4):1758–72.

ANNEX 1: Ex-post power calculation

In this revised method of ex-post power calculations the following points are taken into account:

- Power estimates in this method are test specific. The parameters that were used to calculate power were drawn from the tests of significance that gave the proportions and p-values
- The tests for which the power is calculated are respondent level matched-pair proportion tests, which is the actual study design
- Since the ex-post population variance estimates for different indicators are not necessarily the same, the power estimate are calculated for all indicators. If the ex-post population variance of some indicator other than penta3 is too large, then it may give lower power than the penta3, in spite of its larger sample size.
- Power estimates were calculated for all indicators for all rounds: baseline, 2015 and the DiD analysis, since they are separate tests each of its own merit. Calculating power for 2015 tests would not suffice for the DiD tests, which are important in terms of detecting the effect of the RBF.

The statistical power of a test of equality of proportion is the probability of rightfully rejecting the null hypothesis. It is given by the following formula:

$$\phi \approx 1 - \beta = \Phi(\frac{\Delta}{se} - \Phi^{-1}(1 - {}^{\alpha}/_{2}))$$

Where,
se = standard error of the difference of the two means or proportions
α = significance level = 0.05
Δ = detectable effect size
Φ = standard normal cumulative distribution

In case that the survey is a multilevel cluster one, the standard error in the above formula is an inflated one – that is, design effect is taken into account in it. This is where multilevel cluster survey differs from a simple random sampling, while the rest of the procedure is the same for all. In the following sections we calculated such inflated standard errors for all test and all rounds.

Calculating standard errors for single round intervention-comparison group tests

We use survey design informed matched-pair tests of proportions to calculate proportions and standard errors and p-values. Since the survey design are taken account in these tests, the standard errors given by them are inflated ones rather than the SRS ones. Since the null hypothesis is that the difference between two proportions (treatment and comparison group) is zero, the standard error we need for the power formula is the standard error of the difference between proportions in a matched pair scenario. Therefore, if we have:
p1 = proportion treatment
p0 = proportion comparison group
d = difference between two proportions = p1-p0
Then, the variance for the **d** is calculated in the following way:

$$Var(d) = Var(p1) + Var(p0) - 2Cov(p1, p0)$$

The values on the right of the equation above are given by survey post-estimation variance-covariance matrix.

Calculating Standard Error for the DiD Analyses

In difference-in-difference analysis the difference that is tested is comprised of four terms as follows:

p11 = proportion endline treatment
p10 = proportion endline comparison
p01 = proportion baseline treatment
p00 = proportion baseline comparison
did = difference in difference = p11 – p10 – (p01 – p00)

Therefore, the variance for **did** is calculated in the following way:

$$Var(did) = Var(p11) + Var(p10) + Var(p01) + Var(p00) - 2[Cov(p11, p10) + Cov(p01, p00) + Cov(p11, p01) + Cov(p10, p00) - Cov(p11, p00) - Cov(p10, p01)]$$

The values on the right of the equation above are given by survey post-estimation variance-covariance matrix.

Ex-post Power Estimates for All Matched-pair Tests

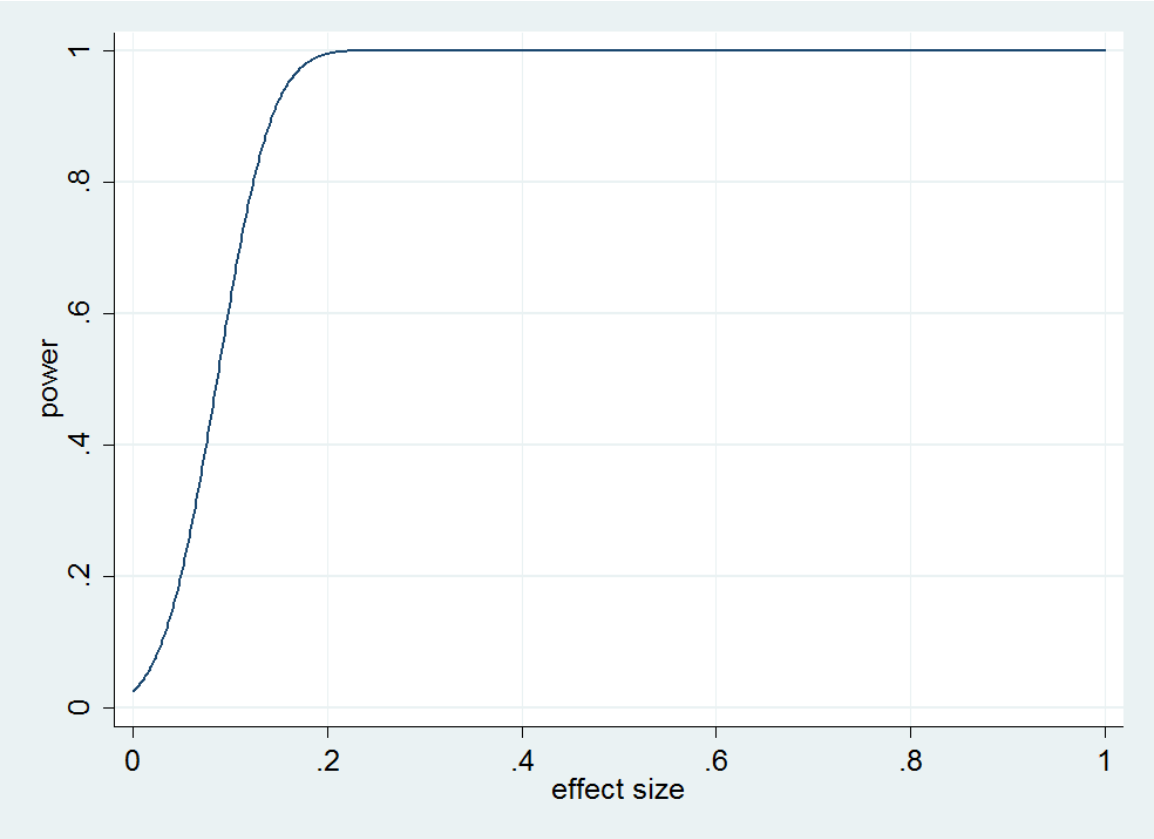
The table below shows that all tests are convincingly powerful in detecting a Δ = .2 (20%) effect size.

Δ =	0.2
α =	0.05
z(α/2) =	1.96

Test	se 2015	power 2015	se 2010	power 2010	se DiD	power DiD
Penta3	**0.0443**	**0.9946**	**0.0431**	**0.9963**	**0.0564**	**0.9438**
Under 5 utilization of OPD	0.0397	0.9990	0.0454	0.9927	0.0510	0.9750
CPR	0.0170	1.0000	0.0106	1.0000	0.0176	1.0000
ANC	0.0332	1.0000	0.0379	0.9995	0.0485	0.9849
SBA	0.0441	0.9949	0.0299	1.0000	0.0469	0.9895
PNC	0.0464	0.9907	0.0316	1.0000	0.0517	0.9719
Institutional Deliveries	0.0417	0.9977	0.0318	1.0000	0.0447	0.9940
Adult morbidity	0.0092	1.0000	0.0157	1.0000	0.0176	1.0000
Adult utilization of OPD	0.0398	0.9989	0.0335	1.0000	0.0416	0.9978

Power Graph for Penta3 2015

Below is the graph of power of Penta3 test for 2015 as a function of effect size, as a demonstration:



ANNEX 2: Results using the original approach

Analysis approach

In Annex 2, we present the original analysis approach by JHU and justification for our changes. We replicated the original approach based on existing analysis plans and Stata do-files. We then aimed to correct or improve the previous analysis as appropriate. We describe the original analysis approach first followed by a revised analysis approach (which is presented in the body of this report).

Original approach

Calculation of estimates for all indicators was conducted by two independent analysts, using standard definitions from Demographic and Health Survey (DHS) and Multi Indicator Cluster Survey (MICS). All data analyses were performed using Stata version 12.0 (13).

Sample weights were calculated for each group to account for selection probabilities. Clustering of observations at the village level, to adjust for within cluster levels of homogeneity (or the intra-cluster correlation), was taken into account by using the Taylor-linearization method in Stata (13). Additionally, strata for application of sampling weights were specified based on facility codes since treatment was randomized at this level.

Exploratory data analysis was conducted to check for duplicate household and respondent ID codes, consistency of data between the household and individual questionnaires, presence of missing, unusual and unacceptable values. At baseline, missing values for indicators such as use of modern family planning methods, antenatal care services, skilled birth attendance, or postnatal care was treated as "none" and this was also done at endline. The same methods for developing indicators were used for the endline analyses. In the analyses represented in this reported, missingness was usually less than 1%. Similar approaches were used when developing child health indicators.

A wealth index score was constructed for each household using principal component analysis of household assets, sources of income, housing characteristics and water source. The method and the questionnaire used for collecting data on wealth have been used in the Monitoring and Evaluation Technical Assistance for Strengthening Health Activities for the Rural Poor (METASHARP) project in the calculation of indicators for the Balanced Scorecard (BSC).

Equity of care indices were estimated by constructing concentration indices. The index was estimated for use of outpatient health services, institutional deliveries, and self-reported morbidity. The calculation followed the procedures described in O'Donnell et al (2008) for micro-data, that is, when values for health variable and wealth measures are available for each individual observation (15).

Following the original design of the study, facility-level summaries were used to estimate the Wilcoxon matched-pairs signed-ranks test for each indicator, comparing treatment and comparison groups. To account for time trends, we used a difference-in-difference (DiD) approach for estimating the effect of the intervention while accounting for unrelated time trends. The analysis was completed at the facility-level for comparing treatment and comparison groups at baseline and endline. The DiD analysis was conducted for all indicators by use t-tests. The concentration index however is not directly measured but it is a measure of inequality that is estimated as the transformation of a variable of interest on fractional rank of wealth for a population. It ranges from -1 to 1, i.e. from a perfect pro-poor distribution to a pro-rich distribution. A mathematical

operation on such a variable within the context of a difference-in-difference model does not lend itself to the same interpretation as seen when dealing with directly measured variables. All tests for equity indicators are done on the population proportions, rather the concentration indices. Mean proportions and their standard errors are reported for all indicators with their related baselined or endline matched-pair sign-rank test p-value, and DiD difference and test p-value. Concentration indices and their confidence intervals are presented for the equity indicators.

Revised approach justification

A midterm evaluation was conducted in 2012 by JHU. We were tasked with conducting the endline household survey and finalizing the impact evaluation for the RBF intervention. Our original task was to replicate the original analysis implemented by JHU in 2012. After conducting the original analysis as planned, we identified some areas for correction and improvement. This section outlines any changes or alterations made to the original analysis, specifically on the indicator definitions, sample weights, and testing design. The revised analysis corrects for basic errors in the analysis plans to false assumptions and disconnects between the analysis and strength of study design. These revisions produce more precise estimates and test p-values.

Changes to indicator definitions: Table 3 describes changes we proposed to the original indicator definitions. Although at baseline, indicators were defined by the standard global DHS and MICS definitions, for the purpose of the RBF evaluation, we are interested only in services delivered at public facilities. Furthermore, community health workers (CHWs) are not considered skilled providers by the MoPH and should be excluded as a possible skilled provider. Other errors were updated such as excluding health post and traditional birth attendants and possible options for seeking child outpatient care services.

Table 2.1 Changes to indicator definitions

2010 baseline	2015 endline changes
Included CHWs as "skilled provider" for ANC and PNC	Exclude CHWs in definition • Justification: While CHWs provide some services under ANC/PNC, they are not considered skilled providers globally and in Afghanistan.
Included private sector for definition of health facility types for adult outpatient utilization of care	
Included private sector for definition of health facility types for institutional delivery	Only include public sector for facility type
Included all health facility types (public and private clinics/hospitals, health post, and TBA) for child outpatient utilization of care	• Justification: We are measuring the impact of RBF intervention which is only implemented in public health facilities (BPHS and EPHS). Including private sector will dilute the true effect on public health facilities.

Correcting survey sample weights and survey design analysis: The weights calculated at baseline did not adequately represent the sampling design. The first stage of sampling health facilities was stratifying by health facility type and randomly selecting matched-pair health facilities. Then for each health facility, two clusters were sampled, and then in each cluster, 24 households were sampled. This requires considering the probability of sampling a matched-pair, not a single health facility within the strata. The previous weights and Stata code did not reflect this design. We also set the survey design analysis in Stata to account for the entire design of the household survey.

Approaches for testing the intervention: Previously, "matched-pair analysis" was conducted using the Wilcoxon sign-rank test, however, this test was run on facility level averages, ignoring the survey design of multi-level sampling and were not weighted. While they argued that the main assumptions of the t-test are likely to be violated due to the assumption that cluster-specific proportions are normally distributed and have equal variances, we are not testing the proportions but rather their differences, which are expected to be normally distributed. This is because under the null hypothesis, the difference between the two groups is due to random error. Therefore, we assume the difference is normally distributed. The normal distributions was confirmed through histograms of the differences in proportions. Therefore, for the matched-pair facility level analysis, we use Chi-square tests and include the sample weights in the analysis. At end-line, we have 67 matched-pair health facilities. Additionally, because we included the survey design in the analysis, it is possible to have matched-paired analysis at the household level, which would appropriately reflect the power of the study and the test p-values. At endline, we have 6,523 matched-pair households. We also provide a before and after analysis of baseline and endline. We set the analysis for DiD as a panel dataset.

Original analysis results

Both intervention groups made improvements on all RBF health indicators since baseline. Large improvements are particularly seen for coverage of PENTA 3 vaccination and skilled birth attendance. At endline, the treatment group shows higher coverage of all indicators compared to the comparison group except for modern contraceptive use. However, none of these differences are statistically significant. The difference in difference analysis shows no positive effect of the RBF intervention. Only postnatal checkup had a positive change but this was not statistically significant.

Table 2.2 RBF indicators, original approach

Indicator	**Baseline (Wilcoxon signed-rank test)**			**Endline (Wilcoxon signed-rank test)**			**DiD (t-test)**	
	Treatment Mean (SE)	Comparison Mean (SE)	p-value	Treatment Mean (SE)	Comparison Mean (SE)	p-value	Diff.	p-value
Current use of modern contraceptives	10.2 (1.0)	10.5 (1.0)	0.902	13.1 (1.2)	15.0 (1.4)	0.516	-1.5	0.234
At least one antenatal checkup from a skilled provider	57.9 (2.9)	57.3 (3.1)	0.993	74.3 (2.3)	73.6 (2.8)	0.878	-0.8	0.818
Skilled birth attendant present at last delivery	28.1 (2.2)	22.9 (2.0)	<0.05	48.1 (3.1)	42.3 (3.1)	<0.10	-0.4	0.912
Postnatal checkup within 42 days of delivery by a skilled provider	26.4 (2.5)	25.0 (2.2)	0.494	41.0 (2.6)	35.8 (2.6)	0.116	2.8	0.497
Children (12-23 months) received Pentavalent 3 vaccination	34.8 (2.5)	31.2 (2.7)	0.206	72.9 (2.7)	74.6 (2.5)	0.666	-1.9	0.740

The equity indicators presented in Table 6 show the mean proportions and signed-rank test p-value with the difference in difference test on the mean. The concentration indices are also presented. Institutional delivery was statistically different between intervention groups at baseline and at endline, but the DiD analysis does not show significance in the change. Child utilization of outpatient services was also statistically significant at baseline but no longer at endline. At baseline, the treatment group was more pro-poor (closer to zero) but at endline, the comparison group is more pro-poor (more negative) across all indicators except for adult morbidity. The change for adult morbidity is positive and this is statistically significant ($p<0.05$).

Table 2.3 RBF equity indicators, original approach								
Indicator	**Baseline (Wilcoxon signed-rank test)**			**Endline (Wilcoxon signed-rank test)**			**DiD (t-test)**	
	Treatment	Comparison		Treatment	Comparison			
	Mean (SE) Equity score (CI)	Mean (SE) Equity score (CI)	p-value (mean)	Mean (SE) Equity score (CI)	Mean (SE) Equity score (CI)	p-value (mean)	Diff.	p-value (mean)
Institutional delivery	26.3 (2.1) 0.134 (0.058,0.210)	21.5 (2.0) 0.169 (0.086,0.252)	0.073*	46.6 (2.9) 0.127 (0.066,0.189)	41.6 (3.0) 0.119 (0.049,0.189)	0.066*	-0.7	0.834
Child (< 5 years) utilization of outpatient services	70.0 (2.3) 0.028 (-0.006,0.062)	76.1 (2.3) 0.031 (0.012,0.050)	0.013**	61.1 (2.7) -0.002 (-0.047,0.043)	60.3 (2.8) -0.038 (-0.073,-0.004)	0.227	5.6	0.195
Adult (> 5 years) morbidity	22.7 (1.2) 0.009 (-0.050,0.068)	25.0 (1.2) 0.013 (-0.035,0.061)	0.12	12.6 (0.7) -0.041 (-0.090,0.008)	12.4 (0.6) -0.009 (-0.053,0.34)	0.667	3.2	0.069
Adult (> 5 years) utilization of outpatient services	73.1 (1.9) 0.007 (-0.038,0.051)	77.4 (1.7) 0.009 (-0.023,0.040)	0.116	80.5 (2.0) 0.052 (0.025,0.079)	79.9 (2.1) 0.008 (-0.014,0.031)	0.648	4.3	0.146

We conducted a before and after analysis by removing the treatment design to test whether the effect of improvements over time. All indicators show statistically significant positive changes in the before and after analysis. Child utilization of outpatient services has declined since baseline, which can be due to less sick children or changes in care seeking practices. This should be further analyzed. On the other hand, while adult morbidity has declined since baseline, adult utilization of outpatient services has increased. The most dramatic improvement is in the coverage of PENTA 3 vaccination by an absolute increase of 40.2 percent.

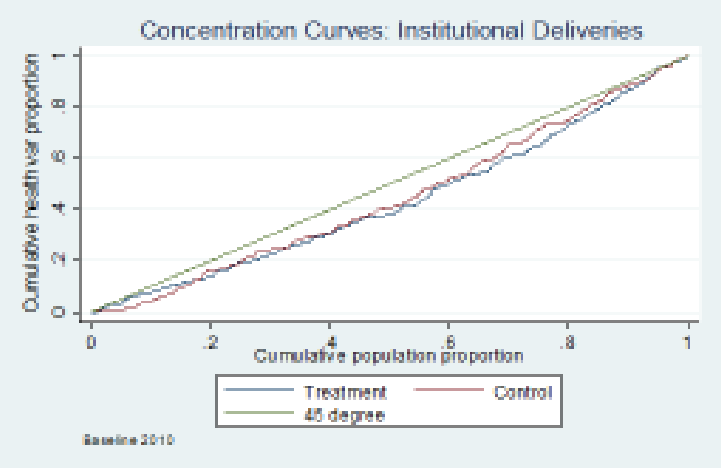
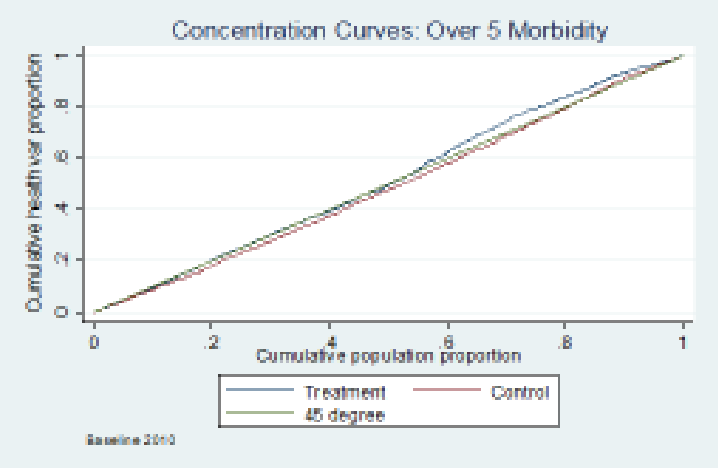
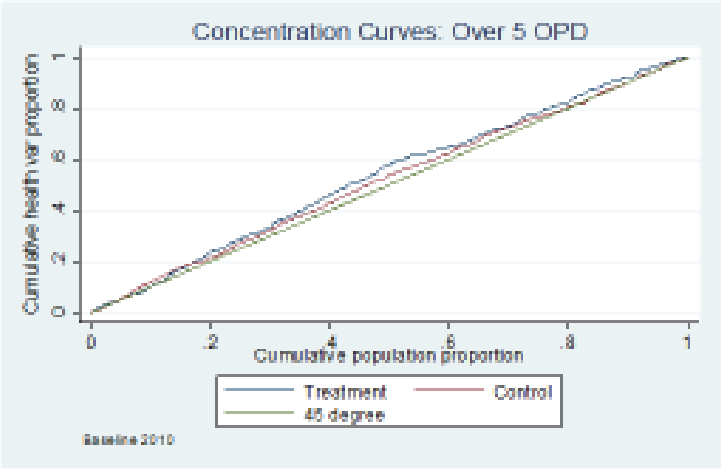
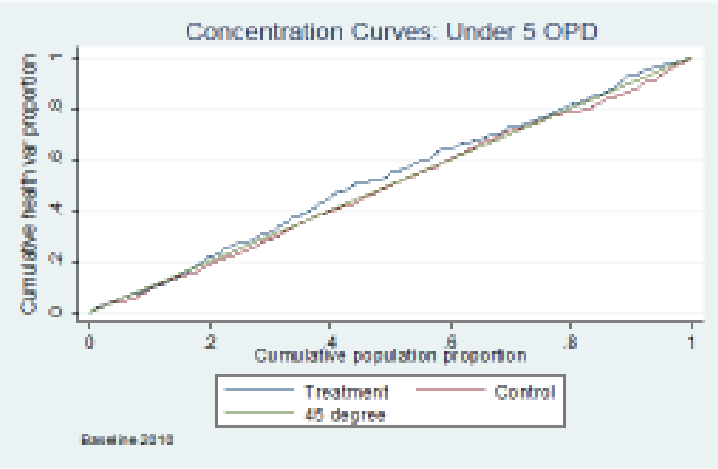
Table 2.4 RBF indicators, original approach, before and after test				
Indicator	2010 (Before)	2015 (After)	Difference	p-value
	Mean (SE)	Mean (SE)		
Current use of modern contraceptives	10.4 (0.7)	14.1 (0.9)	3.8	<0.01
At least one antenatal checkup from a skilled provider	58.1 (2.2)	73.8 (1.8)	15.7	<0.01
Skilled birth attendant present at last delivery	26.0 (1.6)	45.2 (2.2)	19.2	<0.01
Postnatal checkup within 42 days of delivery by a skilled provider	26.3 (1.7)	37.9 (1.8)	11.6	<0.01
Children (12-23 months) received Pentavalent 3 vaccination	33.3 (1.8)	73.5 (1.8)	40.2	<0.01
Institutional delivery	24.4 (1.5)	44.1 (2.1)	19.7	<0.01
Child (< 5 years) utilization of outpatient services	73.3 (1.6)	60.8 (1.9)	-12.5	<0.01
Adult (> 5 years) morbidity	23.7 (0.9)	12.5 (0.5)	-11.1	<0.01
Adult (> 5 years) utilization of outpatient services	75.5 (1.3)	80 (1.5)	4.6	<0.01

ANNEX 3: Characteristics of dropped clusters

Table 3.1 provides selected household and resident characteristics for the baseline sample in 2010, the baseline sample without the dropped clusters, and the dropped clusters. The sampling power was determined at the household level and given the dropped clusters, the coverage sample had less than a five percent nonresponse rate. The table shows that the percent of ever educated women remains statistically significantly different for both study groups under all three scenarios. Across most characteristics, there are none to very little differences between the two study groups in the dropped clusters. Wealth status is significantly different at baseline and in the dropped clusters, but removing these clusters results in no difference in the larger sample. We do not believe the dropped clusters could drive any results since they are similar in characteristic to the baseline sample in 2010, and our sample size is sufficiently powered.

Table 3.1 Selected household and resident characteristics*

	Baseline 2010			Baseline 2010 - without 13 clusters			Dropped clusters		
	Treatment	Comparison	p-value	Treatment	Comparison	p-value	Treatment	Comparison	p-value
Mean number of residents per household (n)	8.1	7.9	0.071	8.1	7.9	0.13	8.7	7.9	0.152
Total number of children 5 years or more and adults (n)	23,394	22,915	0.133	22,372	21,871	0.105	1,022	1,044	0.684
Percent of ever educated women (%)	7.0%	5.4%	<0.01	7	5.6	<0.01	6.3%	2.2%	<0.05
Mean age of women (years [SD])	31.4 (9.2)	31.4 (9.1)	0.946	31.5 (9.2)	31.5 (9.0)	0.774	30.9 (8.9)	30.2 (9.5)	0.39
Ever-married women of reproductive age (15-49) interviewed (n)	4,167	4,117	0.017	3,975	3,910	<0.05	192	207	0.751
Ever-married women of reproductive age with a live birth <24 months prior to survey (n)	1,981	1,977	0.314	1,484	1,478	0.356	98	109	0.822
Number of eligible children (<5 years of age) (n)	4,394	4,304	0.687	4,173	4,075	0.871	221	229	0.246
Children 12-23 months (n)	708	683	0.756	671	643	0.709	37	40	0.838
Travel time nearest health facility (< 2 hours)	2291	2461	0.072	2,188	2,370	<0.05	103	91	0.501
Wealth quintile			<0.05			0.435			<0.01
Lowest	20.7	19.4		21.3	18.1		4.2	40	
Second	18.1	21.9		18.9	21.4		11.1	21.3	
Middle	18.6	21.4		18.1	22.8		23.6	12.1	
Fourth	20.0	20.1		18.8	19.1		37.5	15	
Highest	22.7	17.3		22.9	18.6		23.6	10.6	

*p-values are derived from t-tests or chi-square tests

ANNEX 4: Baseline concentration curves