Report Number: ICRR0021974

## 1. Project Data

| | |
|---|---|
| **Project ID**<br>P120435 | **Project Name**<br>RESULTS-BASED HEALTH |
| **Country**<br>Kyrgyz Republic | **Practice Area(Lead)**<br>Health, Nutrition & Population |

| L/C/TF Number(s) | Closing Date (Original) | Total Project Cost (USD) |
|---|---|---|
| TF-13310 | 30-Jun-2017 | 10,986,864.30 |
| **Bank Approval Date** | **Closing Date (Actual)** | |
| 18-Apr-2013 | 31-May-2019 | |

| | IBRD/IDA (USD) | Grants (USD) |
|---|---|---|
| Original Commitment | 11,000,000.00 | 11,000,000.00 |
| Revised Commitment | 11,000,000.00 | 10,986,864.30 |
| Actual | 10,986,864.30 | 10,986,864.30 |

| Prepared by | Reviewed by | ICR Review Coordinator | Group |
|---|---|---|---|
| Judyth L. Twigg | Salim J. Habayeb | Joy Behrens | IEGHC (Unit 2) |

## 2. Project Objectives and Components

### a. Objectives

According to the Grant Agreement (p. 5), the project's objectives were "to pilot performance-based payments and enhanced supervision for quality of maternal and neonatal care in randomly selected rayon hospitals; and to strengthen the Recipient's and healthcare providers' capacity in performance-based contracting and monitoring and evaluating for results." The statements of the objectives were identical in the Project Appraisal Document and ICR.

### b. Were the project objectives/key associated outcome targets revised during implementation?

No

### c. Will a split evaluation be undertaken?

No

### d. Components

The project contained two components:

1. Pilot performance-based payments and enhanced supervision for quality of care (appraisal: US$9.6 million; actual: US$9.57 million). This component was to finance a randomized controlled trial that aimed to assess two alternatives for improving the quality of secondary care at the rayon hospital level: (1) an enhanced supervision scheme using a balanced scorecard (BSC) to assess quality at the facility level, together with a performance-based payment made against facility performance on a quarterly basis; and (2) an enhanced supervision scheme using a BSC alone without performance-based payments. The project was to finance the performance-based payments and operating costs for the Mandatory Health Insurance Fund (MHIF) to manage contracts with health facilities. Data were to be collected from three groups -- 20 rayon-level hospitals operating under each alternative, and 20 hospitals as a control group -- and analyzed by an independent impact evaluation. The component was also to finance a pay-for-performance pre-pilot at the primary health care (PHC) level, beginning in the second half of the project's implementation period, to leverage the lessons and experience of the hospital pilot, and to inform the design of a PHC pay-for-performance scheme in the future.

2. Strengthen the government's and providers' capacity in performance-based payment reform and monitoring and evaluation (M&E) for results (appraisal: US$1.4 million; actual US$1.41 million). This component was to finance training and other technical assistance to the Ministry of Health (MOH), MHIF, and providers to develop their capacity in performance-based contracting and to conduct peer review and self-monitoring and evaluation for results. Support was to be specifically aimed at increasing effectiveness and efficiency in purchasing, regulation, and quality control of health care services, and in M&E. The component was also to support project coordination, management, and monitoring.

### e. Comments on Project Cost, Financing, Borrower Contribution, and Dates

The project was to be financed by a US$11 million grant from the Multi-Donor Health Results Innovation Trust Fund managed by the Bank. The trust fund was also to make US$1 million available for an impact evaluation of the hospital pilot, not included in total project costs. Actual financing and total costs were US$10.98 million. No borrower contribution was expected or made.

The project was approved on April 18, 2013; became effective on July 29, 2014; underwent a mid-term review in September 2016; and closed on May 31, 2019, almost two years later than its originally scheduled closing date of June 30, 2017. It was restructured twice:

- In June 2017, the closing date was extended to September 30, 2018, to account for initial delays in project effectiveness, align with the closing of the Second Health and Social Protection (SWAp2) Project on December 31, 2019, and allow more time for potential scaling up of primary care results-based financing (RBF) to more districts. At this restructuring, the results framework was modified to

adjust indicators, including extending the reporting timescale for monitoring reports and peer verification records as well as adding some indicators to measure improvements in health care quality that were expected as a result of performance-based financing.

- In December 2018, the closing date was again extended, to May 31, 2019, to give more time to prepare for takeover of RBF, allow for procurement of computers, allow a further round of RBF-related indicators to be collected for SWAp2, and permit alignment with a new Program for Results (P4R) operation (Primary Health Care Quality Improvement Program, US$37 million, 2019-2025).

## 3. Relevance of Objectives

### Rationale

Despite a series of innovative health reforms beginning in 1995, the Kyrgyz Republic at appraisal was continuing to experience poor maternal and child health and was off track to meet its Millennium Development Goals for maternal, child, and infant mortality. This subpar performance tended to be a consequence of poor quality of care rather than access barriers. Although the country had introduced capitation payments for PHC providers and diagnosis-related groups (DRGs) for hospitals under previous reforms, providers were still not held directly responsible for results. This project's objectives, under the umbrella of the ongoing health SWAp, were highly relevant as part of the effort to improve the quality of maternal and child health services. The focus on rayon-level hospitals was appropriate, as this was where most maternal deaths were occurring (ICR, p. 20).

The objectives were also highly relevant to government and Bank strategy. The government's "Unity, Trust, Creation" program for 2018-2022, the first element of the country's overall Sustainable Development Strategy to 2040, contained an explicit focus on improving health care quality. The Bank's 2011 Interim Strategy Note for the Kyrgyz Republic contained three priorities, two of which -- improving governance and effective public administration and reducing corruption; and increasing social stabilization through social services, community infrastructure, and employment -- were supported by the project. The objectives remained highly relevant to the Bank's Country Partnership Framework being developed at closing (CPF, 2019-2022), which expressed concern that the country's MCH outcomes were lagging behind those of neighboring countries in Central Asia and the Southern Caucasus due to low service quality. This project was intended explicitly to lay the groundwork for a larger P4R project that is a key element of the current CPF.

The ICR (p. 11) argued that the output-oriented nature of the objectives was a minor shortcoming. However, given the nature of the project -- to pilot a new performance-based contracting mechanism, the first of its kind in the region, that may or may not have produced hoped-for improvements in quality of care -- an explicitly outcome-oriented objective may have been premature.

### Rating

High

## 4. Achievement of Objectives (Efficacy)

### OBJECTIVE 1

#### Objective

Pilot performance-based payments and enhanced supervision for quality of maternal and neonatal care in randomly selected rayon hospitals

#### Rationale

Both of the project's objectives are explicitly output-oriented, pitched at an appropriate level for a project intended to pilot a new mechanism of provider incentive and reimbursement. The theory of change held that Bank financing for the pilot scheme, including piloting of the BSC and of results-based financing for performance-based payments, would facilitate its implementation. Over the longer term, the implementation of the pilot was to lead to informed decisions about adjustment and possible scale-up of the scheme, and ultimately improved quality of maternal and neonatal care.

Outputs

The performance-based contracting model was introduced to more than the planned number of maternity and neonatal care hospitals. Contracts were signed with 22 treatment group hospitals, exceeding the target of 20, and were updated annually based on modifications introduced to the BSC. A system of enhanced supervision was similarly introduced. A National Steering Committee, RBF Secretariat, and RBF Technical Team were established as conditions of project effectiveness. The National Steering Committee was meeting twice annually by project closure, meeting the target. The extended RBF team met four times in 2018, exceeding the target of three annual meetings.

Outcomes

All treatment hospitals received quarterly performance-based payments in accordance with their level of performance, as agreed in the performance-based contracts, meeting the target. 22 eligible rayon hospitals received the authorized performance-based payments within two weeks after the MHIF's receipt of the RBF payment into its bank account, achieving the target of 22. This outcome is clearly attributable to the project, as no other concurrent RBF program was in place for maternity services.

42% of quarterly peer verification records for rayon hospitals were submitted to the MHIF within 30 working days after the end of each quarter, achieving the target of 42%. Ten hospitals completed counter-verification in the last round, exceeding the target of eight hospitals doing so every six months.

Although observable health care quality outcomes were not a part of the project's theory of change, the ICR presented evidence that quality of care at the treatment hospitals improved as a result of participation in the pilot. The average quarterly quality (BSC) score of hospitals included in the treatment group (performance-based contracting plus BSC) increased from 9.3% in 2014 to 81.5% in 2018, exceeding the target of 75%.

These scores increased so rapidly after the beginning of project implementation (significant improvement was noted after only six months) that the project, at the 2017 restructuring, added additional indicators to measure further the actual quality improvement at the participating hospitals. The average BSC score for hospitals in the treatment group changed: for red blood cell availability, from 19% in 2014 to 73% in 2018, essentially meeting the target of 75%; for maternity departments, from 5% in 2014 to 72.5% in 2018, not quite meeting the target of 85%; for pediatric/neonatal departments, from 21.3% in 2014 to 80% in 2018, meeting the target. The ICR (p. 12) reported a summary of the impact evaluation's findings on quality gains, noting statistically significant improvements along some dimensions and no change on others. In the latter case, the ICR speculated, baseline scores were already high, and/or the sample size was too small or follow-up time too short to have expected observable change.

The Borrower's ICR (p. 43) reported that the BSC scores for hospitals that adopted the BSC but not performance-based payments also improved over the project's lifetime, from 8.6% in 2014 to 70.8% in 2018. Scores for the control group (no BSC, no performance-based payments) remained unchanged until they were included in the project in 2017, at which point their average scores quickly rose from 12.3% (late 2016) to 66% (2018). The ICR did not discuss the positive impact of the BSC alone in the absence of the RBF mechanism.

### Rating
Substantial

## OBJECTIVE 2
### Objective
Strengthen the Recipient's and healthcare providers' capacity in performance-based contracting and monitoring and evaluating for results

### Rationale
The theory of change held that technical assistance and capacity development for MOH, MHIF, and participating providers would strengthen their ability to implement performance-based contracting and M&E.

Outputs

The project provided technical assistance to strengthen the purchasing capacity of the MHIF in preparing, negotiating, and managing performance-based contracts with providers, to enhance the regulatory and stewardship role of the MOH, and to strengthen quality control of the MOH/MHIF with a focus on M&E capacity building, conducting facility and user surveys, independent verification, and auditing mechanisms. Managers and staff at 40 rayon hospitals were trained on the BCS, verification and counter-verification mechanisms, monitoring arrangements, financial management, and procurement. Information systems were strengthened at all participating rayon hospitals (including those in the control group).

Outcomes

Quality Committees were formed or enhanced at participating hospitals (in both the BSC + RBF group, and the BSC-only group), meeting monthly to address clinical concerns (such as adverse events) and share best practices (ICR, p. 12).

The impact evaluation baseline survey was completed before the start of the project, and the impact evaluation was conducted as planned. MOH and MHIF produced quarterly monitoring reports within three weeks after the end of each quarter, meeting the target. MOH and MHIF have committed to continuing and extending the BSC and RBF, adopting these tools into their standard systems nation-wide.

### Rating
High

## OVERALL EFFICACY
### Rationale
The project successfully and fully implemented the pilot, as planned, collecting sufficient data and performing adequate analysis to inform a decision on scale-up. Targets for implementation of the pilot and development of capacity to do so were met.

### Overall Efficacy Rating

Substantial

## 5. Efficiency
The PAD (p. 12) stated that the nature of the project -- pilot testing alternative schemes for improving quality, and evaluating the effectiveness of different approaches -- made it difficult to estimate a cost-benefit ratio or economic rate of return. Similarly, the project was expected to have minimal fiscal impact, as it was financed entirely by the Bank. The ICR's analysis (Annex 4) also noted that it was not possible to quantify all benefits potentially generated by the RBF, and therefore it focused on the health gains from better primary care for pregnant women. It found a benefit-cost ratio of 2.83 over a 15-year period (2019-2033), using a discount rate of 4.5%; an internal rate of return (IRR) of 29%; and a net present value (NPV) of US$40.64 million. A sensitivity analysis found the lower-bound benefit-cost ratio to be still favorable, at 2.44, with an IRR of 23% and NPV of US$31.97 million.

The ICR (p. 15) reported that there were no accounts of deviation from project design or wastefulness during implementation. There was a fifteen-month delay between project approval and effectiveness due to lengthy parliamentary ratification procedures and Bank fiduciary and procurement reviews, but on balance, this is considered a minor shortcoming. The ICR (p. 19) also noted minor shortcomings in procurement at the hospital

level involving delayed payments to suppliers and failure to take advantages of opportunities for savings through group procurement.

### Efficiency Rating

Substantial

a. If available, enter the Economic Rate of Return (ERR) and/or Financial Rate of Return (FRR) at appraisal and the re-estimated value at evaluation:

| | Rate Available? | Point value (%) | *Coverage/Scope (%) |
|---|---|---|---|
| Appraisal | | 0 | 0<br>☐ Not Applicable |
| ICR Estimate | ✓ | 29.00 | 89.00<br>☐ Not Applicable |

* Refers to percent of total project cost for which ERR/FRR was calculated.

## 6. Outcome

Relevance is rated High, as the project's objectives were highly responsive to country context and to Bank and government strategy. Their output orientation was appropriate to their level of risk, the project's relatively short time frame, and its function as a pilot. One of the objectives was substantially achieved, and the other was highly achieved. The project successfully and fully implemented the pilot, as planned, collecting sufficient data and performing adequate analysis to inform decisions on scale-up. Targets for implementation of the pilot and development of capacity to do so were fully met. Efficiency is rated substantial, based on favorable findings from the economic analysis and evidence of only minor shortcomings in implementation efficiency. These ratings are representative of only minor shortcomings overall, producing an Outcome rating of Satisfactory.

### a. Outcome Rating

Satisfactory

## 7. Risk to Development Outcome

The MOH and MHIF are committed to institutionalizing the project's approach across the sector. The ICR (p. 14) reported that that the BCS and RBF innovations are spreading to other hospital departments in participating hospitals and to non-participating hospitals, and into primary care (with an RBF pilot having been extended to the primary care level during the last 18 months of the project). There is now strong interest in systemic change that would extend RBF to primary care, recognized by SWAp2. The MOH and MHIF decided in May 2018 to dedicate 5% of the health budget to RBF across all health care facilities (except those at the tertiary level), based on the project's findings, and have requested technical assistance

to develop a health care quality index for all types of facilities (ICR, p. 17). The project's focus on quality was reported to have sparked enhanced interest in quality of care in nursing and medical training (ICR, p. 14).

There are, however, technical risks associated with MHIF's capacity to manage continued implementation of BSC and RBF. Specifically, some key MHIF staff reported that there is insufficient familiarity with clinical work to implement the BSC and verification. In response, trainings are already in place, and the BSC is being simplified, reportedly without compromising robustness (ICR, p. 22).

## 8. Assessment of Bank Performance

### a. Quality-at-Entry

While the project may have been ambitious and high-risk, given its context in a post-Soviet system "not primed for performance-based incentives" (ICR, p. 16), it was well positioned with effective input- and activity-based payments systems already in place. A pre-pilot had been completed in one oblast (region). The project drew key lessons from previous projects in the Kyrgyz health sector, as well as RBF health operations in other countries adapted well to the Kyrgyz context (PAD, pp. 8-9): the need to account for unpredictable government budgets (accounted for here by planning for the project to finance the full cost of the pilot); and the need to ensure adequate MOH capacity to oversee and steer the sector (accounted for here by incorporating capacity strengthening into the project). The design of the RBF smartly allowed hospitals significant freedom in how they could use RBF payments. Rather than disrupting the existing payment system (based on DRGs), it added a performance element to it (ICR, p. 20). Risk assessment was thorough (PAD, Annex 4), gauging only country risk as high and overall implementation risk as substantial (due mostly to the high level of country risk). Mitigation measures included ring-fencing the project, including strict procurement and financial management practices, and continuing to support the government's ongoing recovery and stabilization programs.

Preparation featured consultation and dialogue with a wide range of stakeholders, including development partners (with whom cooperation was already strong through two SWAps), professional associations, and village health committees. A web-based data visualization tool on project progress helped build project visibility. However, the ICR (p. 17) noted that there were no mechanisms to directly involve patients, parliamentarians, media, or the public in project preparation. Additionally, the ICR (p. 21) noted that the Working Group during preparation may have been too large, making MHIF feel like "one of many invited experts" rather than the future principal owner and implementer of the RBF and BSC.

#### Quality-at-Entry Rating

Satisfactory

### b. Quality of supervision

The Bank team maintained focus on results throughout implementation, as evidenced in its responsiveness to M&E findings (see Section 9c). The same task team leader was in place throughout the project's timeline. Supervision missions provided useful advice and support, and monitoring documents were

complete and candid. The project was restructured to take into account early evidence of success, adopting the BSC at control group hospitals when its benefits became clear, and to incorporate more outcome-oriented indicators, perhaps later than was appropriate (ICR, p. 21). There was some concern about the adequacy of transition arrangements, with some MHIF staff reporting under-preparedness to take over management of the RBF and BSC (ICR, p. 22).

### Quality of Supervision Rating
Satisfactory

### Overall Bank Performance Rating
Satisfactory

## 9. M&E Design, Implementation, & Utilization

### a. M&E Design
Given the project's explicit purpose as a pilot, its M&E plan was intensive and multi-layered, including routine measurements of implementation of the performance-based payments and BSC, as incorporated into the project monitoring framework, and regular measurement of hospital performance through the BSC. Routine project-level data collection was to be the joint responsibility of the MOH and MHIF. The BSC included two aspects: structural elements important for institutional functioning, and clinical processes. The BSCs were to be implemented through a quarterly peer evaluation process that would trigger the performance-based payments, as well as a biannual counter-verification mechanism based on multi-stage random sampling of a subset of rayon hospitals. A planned impact evaluation (PAD, Annex 6) was to assess improvements in the quality of care at rayon hospitals under the entire RBF package and with the BSCs alone, as well as the cost-effectiveness of the RBF package. Its qualitative elements were designed to gain insight into the effectiveness of RBF specifically in the Kyrgyz setting.

Key outcome indicators were well defined and were accurate and comprehensive measures of achievement of the objectives. Indicators, baselines, and targets were clearly articulated and specified. The PAD's results framework (pp. 18-20) logically connected planned outputs to anticipated outcomes.

### b. M&E Implementation
All elements of the M&E framework were implemented as designed. Interim findings from the impact evaluation were widely disseminated.

### c. M&E Utilization
M&E findings were regularly used to inform mid-course corrections. For example, as implementation proceeded and the main constraint on quality of care evolved from availability of equipment to providers' knowledge, the BSC was adjusted accordingly, adding new elements on clinical skills and clinical management competence. When the success of the BSC approach became evident, it was quickly

adopted at rayon hospitals in the control group. M&E findings and utilization contributed to the institutionalization of BSC and RBF, in both primary care and secondary care (ICR, p. 22).

### M&E Quality Rating

High

## 10. Other Issues

### a. Safeguards

The project did not trigger the environmental safeguard, as it did not finance large rehabilitations or new construction, and was rated Environmental Assessment category C. No other safeguard policies were triggered.

### b. Fiduciary Compliance

Financial management (FM) was rated Moderately Satisfactory in Implementation Status and Results Reports throughout the project. Oversight of the project's FM was the responsibility of the MOH. Transfer of RBF funds used pre-existing channels between the Ministry of Finance and hospitals, upon approval of the MHIF based on verification of performance. Hospitals were subject to regular FM audits, which were reportedly always thorough and timely. Interim Financial Monitoring Reports were of acceptable quality. According to the ICR (p. 20), the final independent audit report of August 2019 found reasonable assurance that the project's financial statements were clean.

Procurement encountered no major shortcomings by the central authorities or hospitals. Some minor weaknesses were noted at the hospital level, mostly related to delays in paying some suppliers. Capacity building measures were implemented to strengthen accounting and internal control functions, though high staff turnover remains a challenge (ICR, p. 19).

### c. Unintended impacts (Positive or Negative)

None reported.

### d. Other

---

## 11. Ratings

| Ratings | ICR | IEG | Reason for Disagreements/Comment |
|---|---|---|---|

| | | |
|---|---|---|
| Outcome | Satisfactory | Satisfactory |
| Bank Performance | Satisfactory | Satisfactory |
| Quality of M&E | High | High |
| Quality of ICR | --- | High |

## 12. Lessons

The ICR (pp. 22-24) contained several insightful lessons, some of which are highlighted here:

**Reforms to incentivize improved performance at the facility level, typically seen in higher-income settings, can work in lower- and middle-income health systems when the necessary groundwork is laid**. In this case, the country had a set of prerequisites -- an experienced single payer, a culture of strong monitoring and evaluation, and strong support from key champions of reform -- that facilitated piloting and scaleup of new processes.

**Potentially controversial reforms may work best when they complement rather than disrupt existing institutional and financial frameworks**. In this case, the performance-based payments were layered on top of broader, previously-implemented reforms of provider reimbursement mechanisms, facilitating clear assessment of their impact and increasing the likelihood of support among stakeholders.

**Thinking early in a project cycle about provisions for mainstreaming a pilot can increase the probability of sustaining outcomes**. In this case, the project could have incorporated more deliberate and focused transition support.

## 13. Assessment Recommended?

Yes

Please Explain

This and related projects in the Kyrgyz health sector have been disproportionately successful when compared with some other projects in the country, and other health reform projects in the region. Although there are ample data specifically on this project's implementation and results, a more intensive study of the factors underlying strong performance could produce valuable lessons about what works, in what contexts, and why. The ICR itself (p. 23) recommended a longer-term, formal evaluation of the project's sustained development impact, which could be usefully paired with assessment of the complementary health SWAp.

## 14. Comments on Quality of ICR

The ICR provided an unusually clear, candid, and straightforward account of the project's activities and outcomes. It was concise and consistent, both internally and with guidelines. Its narratives and findings supported its conclusions. It explicitly addressed issues of data quality and attribution. It supplemented data from the project's results framework with quantitative and qualitative data from the impact evaluation. The lessons were highly insightful, with attention to sustainability of the project's development impact and its utility for other results-based projects globally.

**a. Quality of ICR Rating**
High