



## Challenges in preparing and implementing impact evaluations in-country

### 1. Background

**Context:** Benin is starting to implement a Results-Based Financing (RBF) intervention in 34 districts. While inputs and cultural accessibility to care are not the biggest obstacle to healthcare in Benin, improving the quality of care and ultimately mother and child health indicators, especially for the poorest, remains a major challenge that RBF intends to tackle.

**Intervention:** The intervention consists in combining different measures: increasing management autonomy, and/or lump sum staff bonuses, and/or RBF bonuses. The combination of those measures defines five study arms: (i) Increased autonomy + RBF bonuses, (ii) Increased autonomy + lump sum bonuses, (iii) RBF bonuses only, (iv) Lump sum bonuses only and (v) Status quo (comparison group). RBF can go up to 40% of staff salary, and the payment for indicators targeted to the poorest is higher.

**Data collection:** The team conducted the baseline survey of the impact evaluation (IE) in the second and third quarters of 2011. Baseline data entry has been completed and clean data is to be released.

### 2. Country Experience

The team faced several challenges while preparing and implementing the baseline survey.

**At the design stage:** First, during the preparatory stage, the status quo comparison group was added last minute to the IE design, mainly because the team was concerned that there would be small differences between the four initial study arms. This implied sudden increases in the budget and the time needed to complete the baseline survey. The team revised their timeline and requested additional budget in order to conduct the experiment and the survey as needed. In addition, while the team was preparing to implement the intervention and the IE, other multilateral donors became interested in financing RBF interventions in the country. The integrity of the comparison group got threatened in the early stages of the IE, and the team had to negotiate a delaying in other donors' activities in order to keep the comparison group long enough.

Determining the sampling frame also raised a few questions and discussions within the team. While the team benefited from political endorsement from the Ministry of Health, they did not obtain the authorization from the National Institute of Statistics to use their census data to build the sampling frame. The team resorted to building the sampling frame themselves, and identified eligible households via preparatory field work. This led to a significant increase in costs, which had not been originally included in the terms of reference of the survey firm and not planned in the original IE budget. It also delayed baseline survey field work.

The team also faced a 2-month long discussion on the connection between households and health

facilities: should the team start by surveying households and asking them about the health facility they most frequently used, or should health facilities be surveyed first, and a catchment area defined around each health facility to locate households “attached” to each facility? Since all health facilities were included in the IE survey, no sampling of the facilities was necessary. The team decided to go for the second solution, and defined health facility catchment areas of 5 kilometers in the South and 15 kilometers in the North of the country. For each facility, ten households were surveyed, based on the initial mapping of sampling units conducted that gave the number of households in each sampling unit.

***During implementation:*** The team launched a competitive bidding for survey firms to conduct both household and health facility surveys. The rationale behind having one firm for both surveys was to save on transaction and transportation costs, especially given the unexpected increases in costs and delays that had already occurred. The winning firm was a local firm, chosen on the grounds of lower costs, but also with the concern of building local capacity. The firm benefited from the presence of skilled staff but it lacked experience in large scale surveys. The World Bank hub team and their partner quality assurance firm Sistemas supported and trained the members of the survey firm during the pilot phase to limit the risks of poor data quality. For example, the length of the household questionnaire was reduced from four hours initially to about half of it, and medical staff were trained to conduct anthropometric measurements and blood testing among households. In addition the team emphasized the supervision of household data collection with the presence of external controllers. As a result, the household data collected was of good quality.

However, the health facility survey was not conducted with such supervision and control. This got combined with very technical content: enumerators had to administer fifteen health facility instruments such as the measurements of infrastructures, motivation, skills, knowledge, absenteeism and direct clinical observation. The measurement of absenteeism also implied two visits, one announced and one unannounced, which added to the complexity – and the cost – of the survey. Finally, the firm did not have a lot of expertise in health facility surveys in the first place. For those reasons, the team is concerned that the quality of the health facility data may not be as good as that of the household data, and is considering recollecting part of the data at follow-up if necessary.

Regarding data entry, again the team highly benefited from the training of the quality assurance firm (Sistemas). In addition, the presence of a highly skilled data manager within the survey firm ensured no major issue arose during data entry, and data entry operators could turn to the local data manager if they encountered issues with the data entry mask, even without the presence of the data quality assurance firm.

### **3. Lessons Learned**

- The time and budget needed for an impact evaluation survey are significant. Teams should be ready to face costs and timing adjustments, which may occur if the design of the experiment is modified, or if the implementation of the sampling frame or of the survey itself is more challenging than expected.
- Political dialogue and decisions play a significant role in the design of the IE and the rollout of the IE survey. Teams should ensure the integrity of the IE design and implementation are maintained, while staying in line with national priorities and political will.

- Household and health facility surveys require a mix of similar and different skills. Teams should conduct two different biddings for each survey in order to maximize the quality of the data.
- Teams should be very wary of low survey firm costs. They should review proposals carefully, and make sure the terms of reference of the survey firm, and the associated budget, include all the tasks the survey firm is expected to perform. In particular, activities pre- and post-field work (such as identifying sampling units or cleaning data) must be budgeted carefully. Anthropometric measurements or blood testing are costly activities and require the procurement of materials that can be time consuming: this should be taken into consideration when the terms of reference, timing and budget of the survey firms are reviewed.
- Impact evaluations require a big team: survey firm staff including enumerators, supervisors, data entry operators and data entry manager, 1-2 data quality assurance expert(s), 1-2 controller(s), 1-2 data analysis expert(s), along with the core IE team (PI, evaluation coordinator, etc.) and the assistance of the World Bank Hub team when needed. Relying on and developing the different skills of those partners is key to the success of the IE. The costs of involving those partners must be included in the IE budget.
- Task Team Leaders need to commit a significant amount of time to get funding, procure and prepare for the IE.