Developing solutions to reduce poverty and improve lives

# Mastering the challenges of impact evaluation in adaptive programmes: Lessons Learned from Nigeria

Karolin Krause
(Coffey International)
14 December 2015

# 1. Introduction

1. Introduction
2. The ‘gold standard’ of (quasi-) experimental evaluation designs
3. The challenges of applying the ‘gold standard’ to a flexible market development programme
4. Lessons learned from Nigeria
5. Concluding remarks

# 2. The ‘gold standard’ of (quasi-) experimental evaluations

# 2. The ‘gold standard’ of (quasi-) experimental evaluation designs

- **In (quasi-) experimental evaluation designs,** research is conducted with an intervention group receiving programme activities, as well as a control group that is similar but does not receive any intervention.

- Typically, a quantitative survey is conducted at the start and end of the programme to capture changes that have happened to the lives of both groups.

- This approach is particularly rigorous because it helps to single out the specific impact of an intervention over and above what would have happened anyway.

*Therefore, experimental and quasi-experimental evaluation designs are considered the ‘gold standard’ for evaluating development interventions.*

# 2. The ‘gold standard’ of (quasi-) experimental evaluation designs

- But achieving the gold standard of a (quasi-) experimental evaluation is difficult, because such designs require:
  - strong quantitative research skills
  - a clearly defined target population
  - ability to identify a comparison group
  - a clear intervention model with a simple, testable mechanism of change

→*The reality in development programmes rarely looks like this.*

***Especially not in adaptive and iterative programmes.***

# 3. The challenges of applying the ‘gold standard’ to a flexible market development programme in Nigeria

# 3. The challenges of applying the 'gold standard' to a flexible market development programme

coffey

## Coffey is delivering the GEMS4 market development programme in Nigeria:

- Running with DFID funding from 2012 – 2017 in Lagos and Northern states
- Seeks to create 10 000 new jobs and improve 500 000 incomes for the poor in Nigeria's wholesale and retail sector.
- Aims to achieve systemic change *indirectly* by facilitating new partnerships, aligning incentives, and encouraging new practices among market actors.
- Tests ideas through pilots and refines them iteratively, to identify and scale up solutions that work.
- DFID expects GEMS4 to measure its results robustly by using a quasi-experimental evaluation design.

GEMS | Growth & Employment in States
Wholesale and Retail Sector

# 3. The challenges of applying the 'gold standard' to a flexible market development programme

- We initially designed a classic quasi-experimental evaluation for GEMS4's three pilot interventions.
- Conducted a large-scale baseline survey amongst retailers and wholesalers in intervention and comparison markets.
- But after the first year of GEMS4 programming:
  - Retailers and wholesalers in the intervention markets had not adopted the new practices promoted by GEMS4 (e.g. mobile money)
  - The pilot interventions had to be reviewed and revised
  - The geographic focus of the interventions shifted

*Iteration is entirely normal in a programme like this….But it did mean that the initial baseline research lost some of its relevance.*

# 3. The challenges of applying the 'gold standard' to a flexible market development programme

- Throughout GEMS4's lifetime it remained difficult to apply the 'gold standard' of quasi-experimental impact evaluation, because:
  - Intervention mechanisms kept changing in response to new opportunities
  - The 'intervention group' kept shifting
  - Retailers and wholesalers 'self-selected' into those embracing new practices and those not adopting them – for reasons that set them apart.
  - Ideally, there would be a pervasive effect and *no* counterfactual.

→*No clear starting point to conduct baselines, as the programme kept evolving and changing.*

→*Blurry boundaries between intervention and comparison groups made it difficult to use a quasi-experimental design.*

# 4. Lessons learned from Nigeria

# 4. Lesson Learned: Conducting pre-baselines

***The challenge:*** *When is the right time for a baseline? How can we avoid that baselines become obsolete?*

***Lessons learned:***

- Conduct a **qualitative pre-baseline** at the start of any intervention **to test the intervention mechanism** and assumptions about how it should work.
- Our qualitative pre-baselines involve for example:
  - Action Research with retailers and wholesalers
  - Participatory mapping exercises to understand relationships between market actors
  - Qualitative interviews and focus groups with market actors and key informants

Lesson 1: Only invest in a large-scale quantitative baseline once the mechanism of change has been proven.

# 4. Lesson Learned: Mitigating risks

***The challenge:*** *How can we minimise the risk of investing a lot in a baseline sample that might become obsolete if the programme iterates?*

***Lessons learned:***

- Draw a smaller survey sample at baseline, and a larger sample at endline
- Requires a larger sample overall to make up for imbalance but allows to invest in a large sample only if the intervention has actually survived until the endline
- Coping strategies such as this can help manage the risks evaluating open-ended programmes

Lesson 2: Limit the resources invested in the baseline sample and invest more at the endline.

# 4. Lesson Learned: Complementing baseline and endline

***The challenge:*** *How can we measure the programme's impact robustly if we cannot identify strong counterfactuals?*

***Lessons learned:***

- We still conduct baseline and endline surveys in intervention and comparison groups, where possible. But this alone is not enough.
- We complement this with **continuous qualitative research** that is conducted in real time throughout the lifetime of the programme. This involves:
  - Tracking a cohort of beneficiaries to explore how their lives have changed over time
  - Monitoring contextual factors that may influence changes in the market
  - Testing assumptions and linkages underpinning the intervention logic, and
  - Building a narrative around what has or has not been achieved, *how and why*.

Lesson 3: Mitigate the lack of strong counterfactuals by conducting continuous qualitative research and robust before-after comparisons.

# 4. Lesson Learned: Make M&E everybody's responsibility

***The challenge:*** *How can we conduct continuous qualitative research in between baselines and endlines when the M&E team has limited time and resources?*

- M&E initially perceived as a sphere of its own with little support from programme staff in capturing changes and results.

***Lessons learned:***

- Demystify Monitoring and Evaluation
- Empower programme staff to do informal research at the grassroots in real time
- Make monitoring, reflection and learning everybody's task

Lesson 4: Break up the strict divide between the M&E and programming teams and make research and learning everybody's responsibility.

# 5. Concluding remarks

# 5. Concluding remarks

- The ‘gold standard’ of (quasi-) experimental evaluation designs is not necessarily the most appropriate for evaluating adaptive and iterative programmes.
- Given the uncertainty inherent in such programmes we cannot rely on quasi-experimental surveys alone to measure their success.
- Qualitative research is essential for testing mechanisms of change and refining them continuously throughout the lifetime of the programme.
- Monitoring, Evaluation and Learning must be everybody’s responsibility because there is too much to do and too little time in an adaptive programme, to rely on formal M&E processes only.

*Thank you!*

*Any questions?*