



Oxford Policy Management



Challenges of achieving both accountability and learning:

Case study of the evaluation of a payment-by-results programme

Presentation at the UK Evaluation Society Conference

Alex Hurrell
Lucrezia Tincani

May 2017

Structure of this presentation

1. Background

- Overview of programme being evaluated
- Challenges of the evaluation aims

2. Challenges of designing the evaluation...

- for accountability
- for internal learning
- for external lesson sharing

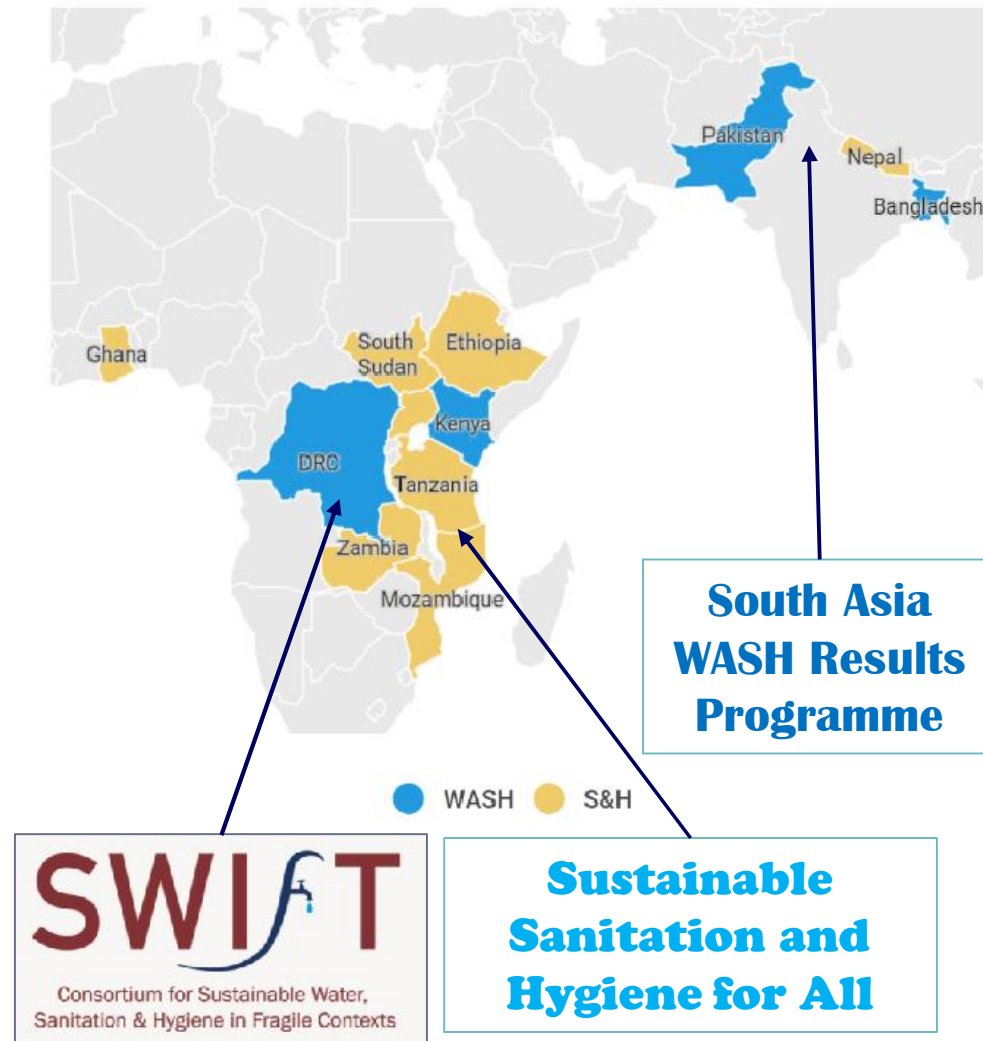
3. Reflections on evaluation design



Background



The WASH Results Programme



- Improving water, sanitation, and hygiene services for 4.5 million people
- Over 4 years (2014 to 2018)
 - Outputs by Dec. 2015 (*MDG deadline*)
 - Outcomes by Mar. 2018
- Delivered by three INGO consortia (Oxfam, Plan, SNV)
- Across 12 countries

It is a payment-by-results programme



Aims of the evaluation

Lessons on

- Under which circumstances were activities **effective**?
- **Efficiency** of activities and of the verification?
- **Relevance** of programme design?
- Likelihood of **sustainable** results?
- Likelihood of equitable **impacts**?



- *For **accountability** on impacts*
- ***Internal learning***
- ***External lesson sharing***



Challenges on evaluation design for different aims



Evaluation = for accountability on outputs/outcomes?

Already covered by the third-party verifier

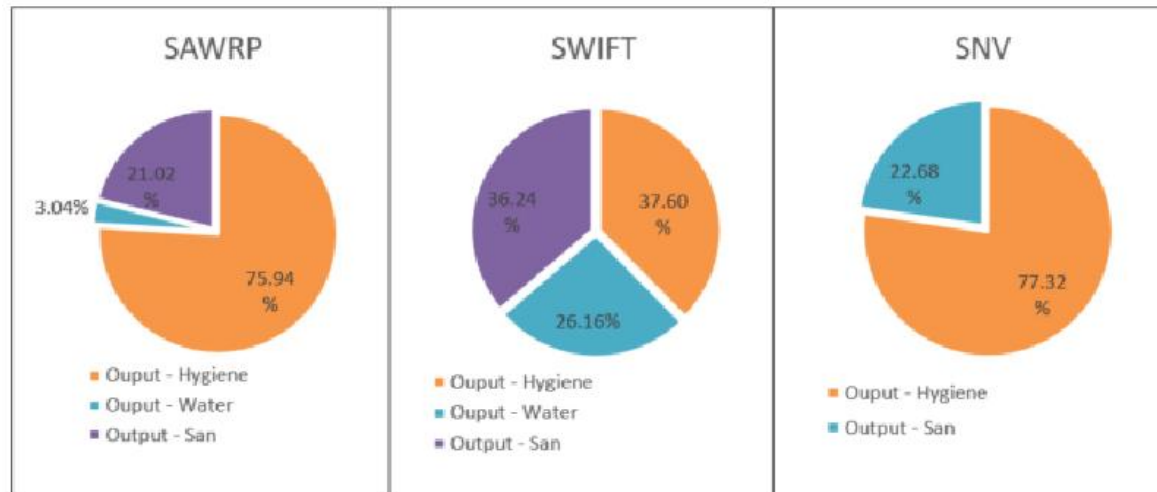
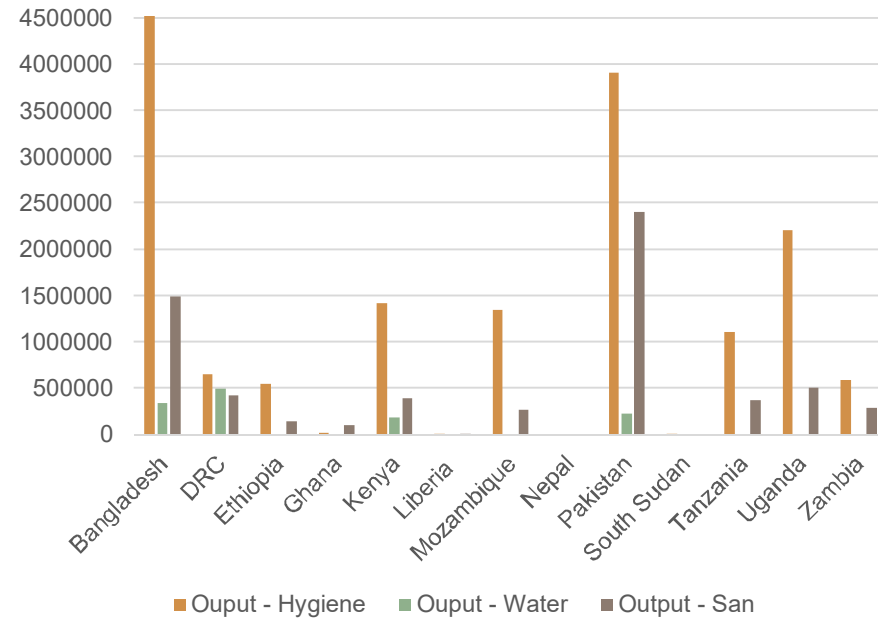
- The verifiers checked if outputs & outcomes were achieved
- The evaluation checked if the verification process was *relevant, efficient and effective*

BUT is there a potential tension of having the Verification Team and the Evaluation Team under the same contract?

Evaluation = for accountability on impacts?

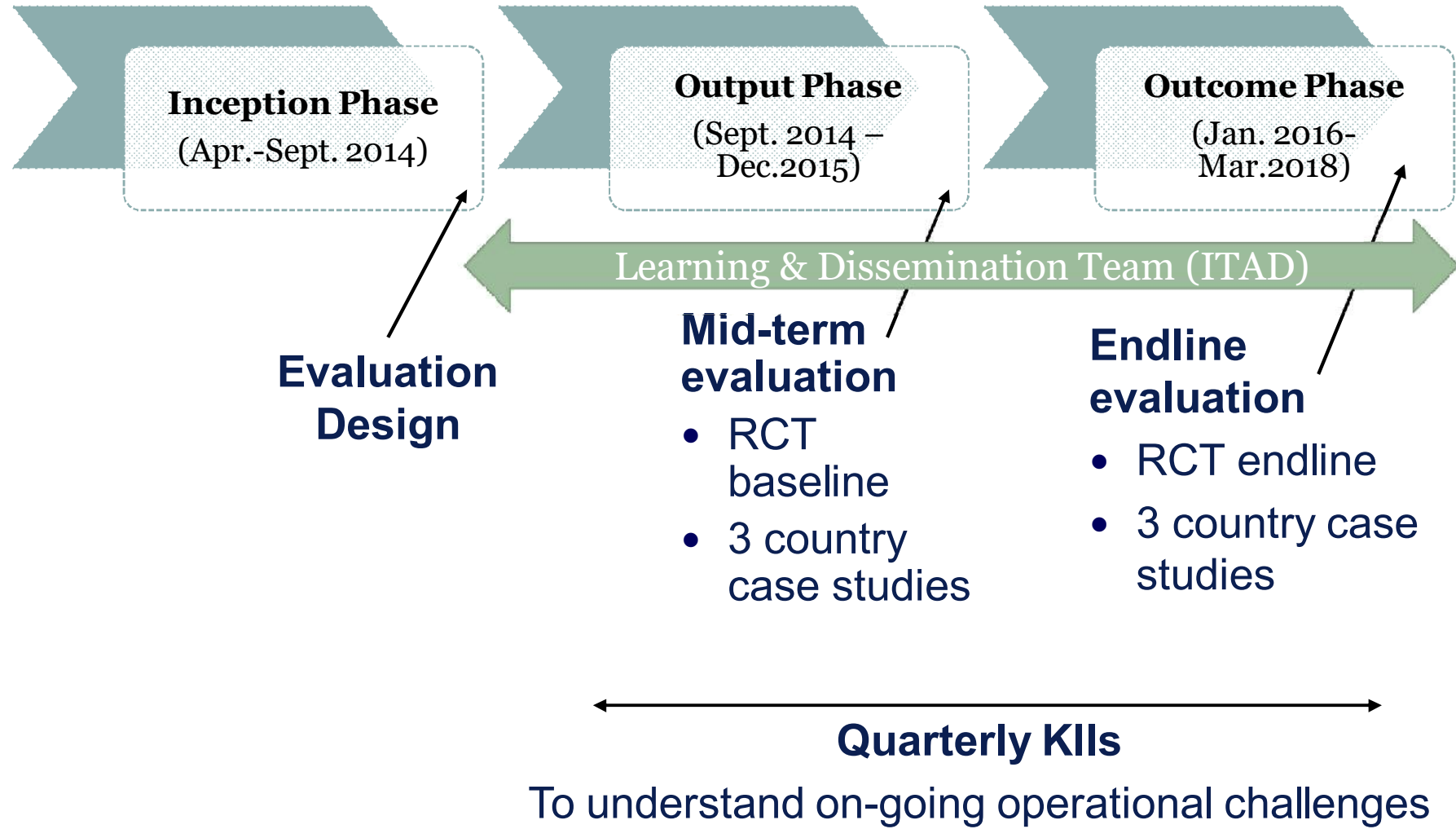
Evaluating if intended impacts were achieved: challenging for such a complex programme

- Very different sub-programmes: different scale of beneficiaries; different WASH activities
- Different verification procedures



Evaluation = for internal learning?

Two rounds of evaluation findings



Evaluation = for internal learning?

Timing of evaluation findings too late to inform NGO activities?

Lessons which could have been useful for internal learning:

- Bonus for overachievement could be used to incentivise performance
- Standardisation of verification indicators across implementing organisations can facilitate the verification exercise
- Having separate verifiers for separate sub-programmes can hinder cross-programme learning
- The PBR mechanism acts best on activities that are within the implementers control (e.g. don't rely on government activities)

→ Too much implementation pressure to put implementation on hold and wait for evaluation to take place

- SNV countries started implementation before a baseline could be done
- PEW case studies pushed into 2016 as “2015 was too busy for implementing organisations’ teams”

Evaluation = for external lesson sharing?

(Some) Lessons for DFID/NGOs doing future PBR programmes:

- Longer inception phase needed to define suitable monitoring framework, reporting process and QA procedures
- Verification indicators need to be designed before implementation starts, so that Consortia can properly resource the monitoring effort.
- Tailoring the intensity of verification process, along the TOC: more light-touch for “process milestones”, more in-depth for Outcomes
- Chose a homogenous set of implementation activities which need a similar verification process (efficiency gains; allows verification at scale)

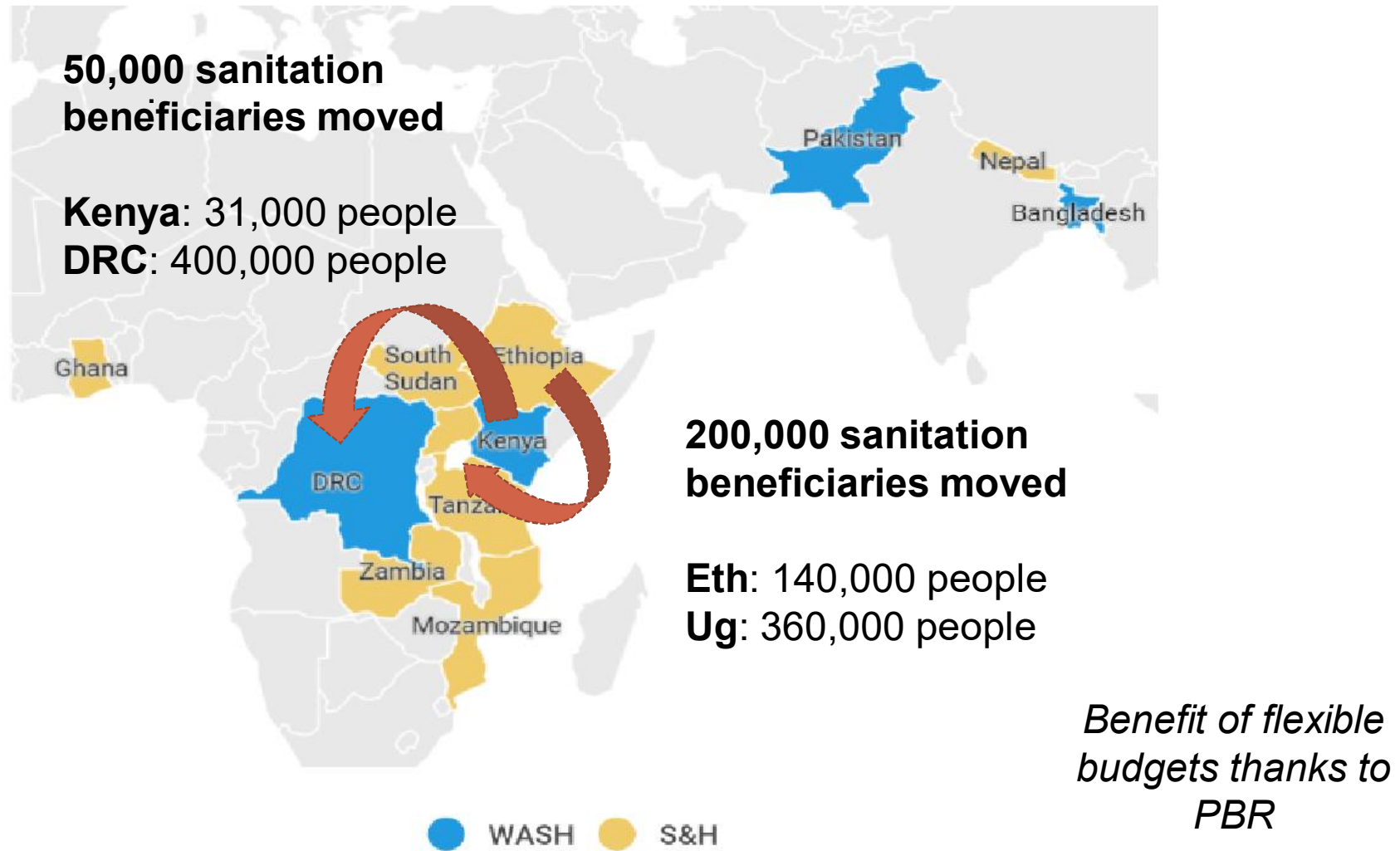
→ Opportunities to use lessons in Phase 2 of the programme

→ Opportunities to inform DFID’s future PBR programming?



**What effect did the PBR
modality have on programme
implementation?**

Results were met, but not always in planned locations



Did PBR strengthened monitoring systems?

In general YES

- SWIFT: More regular (monthly) monitoring requested from partners
- SAWRP: Database cleaned up to avoid double counting
- SSH4A: Opted for mobile monitoring – allows data quality checks and live data

BUT only improved accountability / visibility *if* reporting was linked to government district plans (e.g. SSh4A used districts governments to monitor progress)

But did PBR improve implementation?

In theory, better up-to-date monitoring data would flag up areas lagging behind and allow corrective action

- **Yes?** Focussed the mind on how beneficiaries can be reached
 - SAWRP: physically visited communities more often
 - BBCMA: did FGDs on how best to reach radio listeners
 - SWIFT built durable & reliable water systems (solar panels, strong tubing)
- **No?** Cutting corners to match verification requirements
 - SAWRP: “CLTS light” – spent less time in communities (*sustainability?*)
 - SWIFT: Poor ‘triggering’, simply moved targets when did not reach ODF
 - SWIFT mainly rehabilitated existing water points, reducing risk (*equity?*)
 - SSH4A: rushed BCC (no door-to-door) – unlikely to lead to hand-washing behaviour change

→ **Effects confounded by time pressure of Dec.2015 MDG deadline**

Results achieved – but at what cost?

- Overburdened M&E staff
- SWIFT contributed own funds to complete water systems
 - Cost of electrifying water supply systems
 - Cost of hitting a ‘dry’ borehole
- Local partners did not get paid if did not reach results (e.g. Kenya, Uganda)
- Stretched relationships with government partners
 - E.g. Urban water utility in Nairobi



Reflections on accountability vs learning

A few questions for discussion...

1. Is it appropriate for PBR verification (*hard accountability*) to come under an evaluation contract?
2. In general, perhaps more helpful to think of LT vs ST, or programme- vs sector-level learning - rather than accountability vs learning?
3. If so, is this actually an issue of scope and focus, rather than an inherent trade-off?
4. But, how to mitigate the risk that taking on an ongoing "learning partner" role can undermine evaluation independence?



Oxford Policy Management

Thank you