DFAT Review of operational evaluations 2014

Discussant: Terence Wood (terence.wood@anu.edu.au)
The plan:

1. My interpretation of findings
2. My findings about how they found their findings
What follows is definitely not an attack!
Key findings (as I see them):

Most DFAT evaluations are quite good
Evaluations being thought about intelligently in DFAT (e.g. this report)

But DFAT does not invest nearly enough time nor people-power into the evaluation process
• average evaluation cost as a proportion of activity value 0.37% - are we really *that* confident that aid works? (DFID 1.9%)
• evaluation attrition
Evaluation attrition

2014 - % of evaluations surveyed (blue); % of evaluations (red)

- ok/good quality: 77%
- useful recs: 66%
- mgt responses found: 46%
- evaln published: 38%
Methodology! Findings about how they found their findings

“I’ve been working out for 6 months but all my gains have been in cognitive function.”
This study is good!

• convincing enough and important points identified
• methods are spelled out (and limitations noted; albeit in an appendix)
• coding/scoring schema made sense

But in future...

• using the population or increase the sample (two years?)
• **Careful with causal claims**
• do evaluations have an impact?
• **in-depth case studies (particularly of failure)**
An example where causal care was needed

No reason to believe “high or very high” quality evaluations are a “high or very high” because they have fewer working days. It could be that many ‘adequate’ evaluations would have been worse with fewer days. Maybe in some cases, more days were needed cause it was a more difficult job. “Maybe high or very high” evaluations were good and quick because you had really good evaluators.