

Department of Health policy pilots: how the purpose of piloting matters for evaluation and learning

Stefanie Ettelt & Nicholas Mays
London School of Hygiene & Tropical Medicine

Department of Health seminar, 5 February 2013

Outline

- The rationale for piloting
- The study
 - Aim and objectives
 - Methods
- Findings
 - A typology of purposes of piloting
 - Three case studies
 - Key issues
- More observations
- Questions for discussion

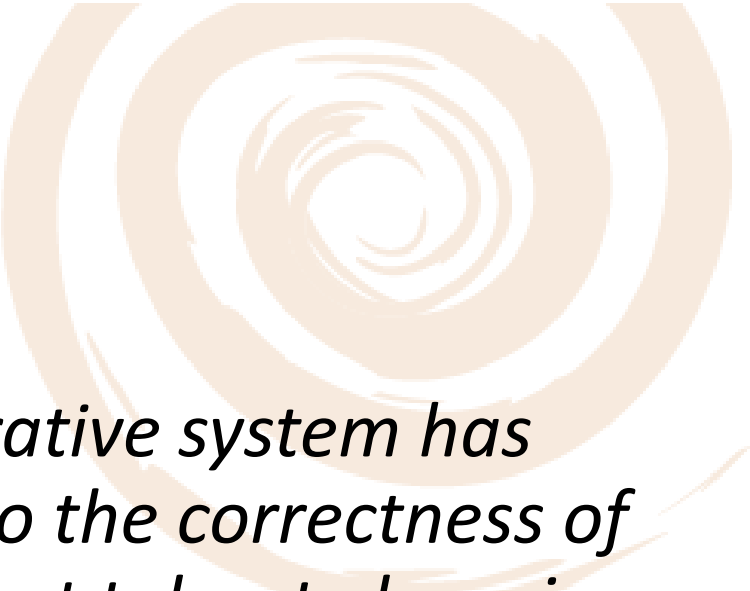


The 'official' rationale for piloting

- Improving policy and policy outcomes by using evidence to inform decision-making
- Piloting as experimentation: Testing in real settings whether and/or how a policy works, prior to roll out

“Rigorous early evaluation of a policy (or some of its elements) before that policy has been rolled out nationally and while [it] is still open to adjustment in the light of the evidence compiled.”

(Cabinet Office, Trying it out, 2003: 11)



“If the political and administrative system has committed itself in advance to the correctness of efficacy of its reforms, it cannot tolerate learning of failure. To be truly scientific we must be able to experiment. We must be able to advocate without that excess of commitment that blinds us to reality testing.”

(Campbell, Reform as experiment, 1969)

Aim and objectives



Aim

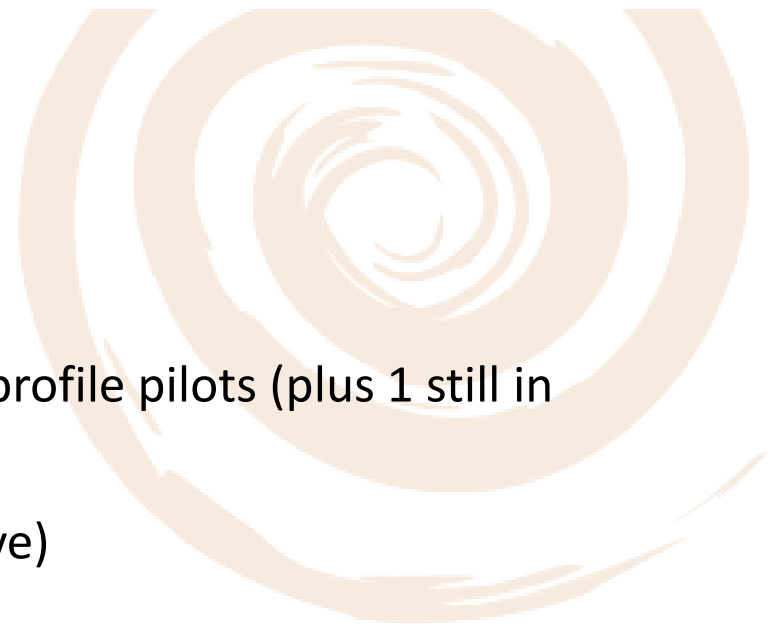
- To explore the nature and purposes of selected DH policy pilots and how these affect their evaluation

Objectives

- To understand the purposes of selected high-profile policy pilots and how they have been implemented
- To analyse how these purposes have influenced decisions about the evaluations of these pilots
- To identify key issues which can inform the design of policy pilots and evaluations in future

Methods

- Multiple case study design
 - Case studies of 3 relatively large, high profile pilots (plus 1 still in progress)
 - Largely retrospective (plus 1 prospective)
- Interviews (n. 30; 60-160 min)
 - DH Officials; leads/managers in pilot sites; evaluators; others
 - Semi-structured; allowing for flexibility (case, role, time of involvement)
- Documents
 - Policy documents (e.g. White Papers); circulars to local agencies; research briefs; discussion papers; response to reviewers
 - Evaluation reports; published papers; commentary
 - Documents produced by pilots sites (e.g. presentations)



The three case studies



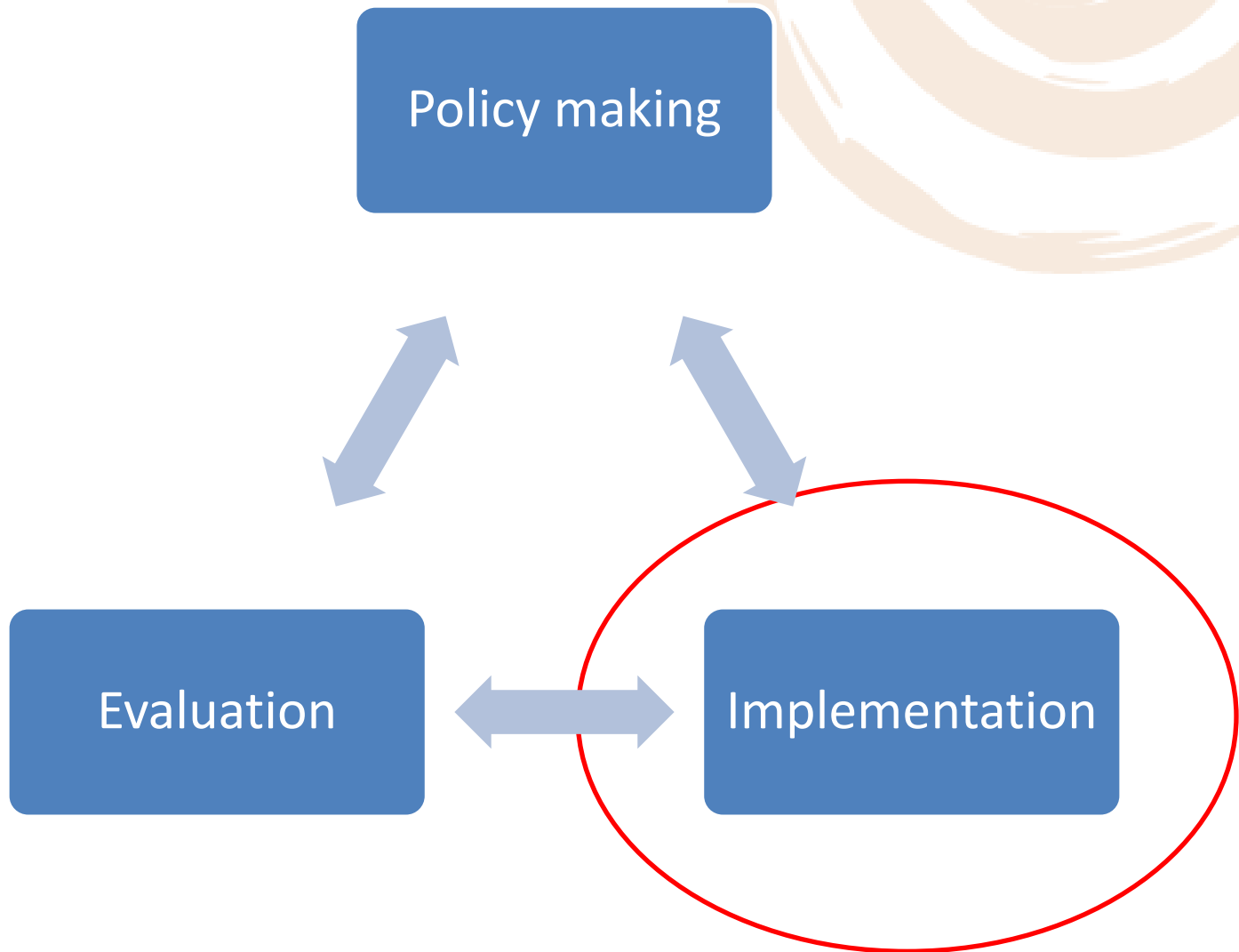
- Partnership for Older People Projects pilots (POPPs)
 - Interventions aimed at keeping older people out of hospital through working in partnership across NHS, local authorities etc.; Programme budget £60m.; 2006-09
- Individual Budgets in Social Care pilots
 - Providing social care users with a budget to enable them to select their own providers and type of care; incl. RCT; Programme Budget £5m.; 2006-08
- Whole System Demonstrators (WSD)
 - Testing of new ways of delivering health and social care services organised around Telehealth and Telecare; incl. RCT; Programme Budget £30m.; 2007-2011

Outline

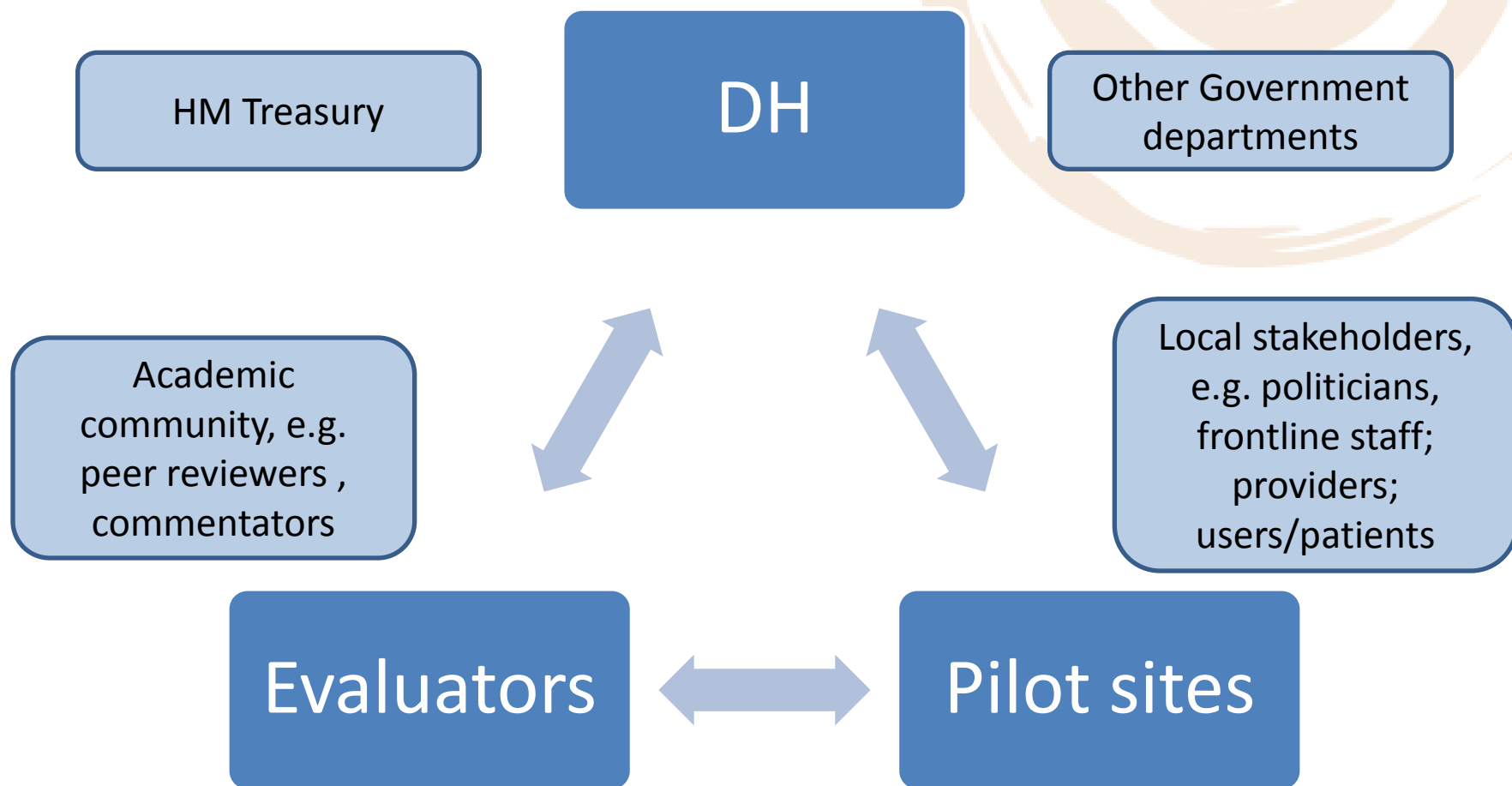
- The rationale for piloting
- The study
 - Aims and objectives
 - Methods
- **Findings**
 - A typology of purposes of piloting
 - Three case studies
 - Key issues
- More observations
- Questions for discussion



The three activities involved in piloting



Main stakeholders in piloting



Typology of the purposes of piloting



Piloting for experimentation

- Testing whether a policy ‘works’ and what outcomes it can produce

Piloting for implementation

- Early implementation
 - Piloting as an opportunity for initiating, and investing in, local change through implementation in pilot sites
- Demonstration
 - Showing others (non-sites) how the policy can be implemented successfully (like a cooking demonstration)
- ‘Pathfinder’/‘Trailblazer’
 - Learning how to operationalise policy, how to overcome implementation barriers and how to improve processes and outcomes

Outline

- The rationale for piloting
- The study
 - Aims and objectives
 - Methods
- Findings
 - A typology of purposes of piloting
 - **Three case studies**
 - Key issues
- More observations
- Questions for discussion



Partnerships for Older People Projects (POPP) pilot

- Initially set out to foster innovation and facilitate learning (pathfinder)
 - Large number of sites and projects: 29 sites, recruited in 2 rounds, 146 'core' projects plus over 500 'other' projects
 - Pilot sites were required to commission local evaluation
 - The initial proposal for evaluation suggested a case study design
- Subsequent shift to emphasise outcome evaluation, suggesting a focus on experimentation
 - Stronger interest in measuring outcomes and costs
 - This changed the expectations of the evaluation while in progress, for example, with respect to randomisation (there was a sub-sample comparison included with data from the British Household Panel Survey)
 - It became very difficult to deal with the diversity of sites and projects.

Individual Budgets in Social Care pilot



- Initially set out as an experiment
 - A smaller number of sites (n=13), representing a diversity of capabilities (including those seen as ‘laggards’)
 - RCT, randomised at user level; in combination with process evaluation
- With characteristics of a pathfinder
 - There was a lack of clarity of what was to be evaluated (‘the model’), corresponding with an initial lack of clarity about what was to be implemented;
 - The ‘In Control’ model became the dominant approach; interviewees disagreed whether this was desirable, with some arguing that an opportunity to test other approaches had been lost
- Irrespective of which the pilot turned into an early implementer
 - A policy decision was made to announce national roll out in October 2006
 - This led to confusion in sites and undermined the approach to evaluation, i.e. the purpose of randomisation became more difficult to “sell”

Whole System Demonstrators



- Was designed as an experiment
 - Sites (n=3), involving 6000 participants according to power calculation
 - ‘Pragmatic’ RCT, with cluster randomisation at practice level, in combination with process evaluation (5 themes)
- Which had a separate early implementer component
 - WSD ‘Action Network’ commissioned to promote telehealth, disseminate existing evidence and distribute best practice experience
- And was also intended to be a ‘demonstrator’
 - As the name suggests
 - Some sites selected based on their previous experience of delivering telehealth and/or telecare; but they could not use the existing case load for the trial, so had trouble recruiting participants

Implications of multiple purposes for experimentation

- Criteria for selecting pilot sites
 - Selecting ‘high achievers’ (for demonstration) may not be representative of the conditions when the pilot becomes national policy (i.e. Threat to external validity)
 - Selecting sites with previous experience: may not be easily compatible with randomisation (i.e. Threat to internal validity), unless existing caseload is excluded (which delays implementation)
- Defining what is to be evaluated
 - Outcome evaluation is difficult (if not impossible) if there is a great diversity of interventions or if there is uncertainty about how an intervention is operationalised (pathfinder)
- If there are local evaluations their relationship to the national evaluation has to be clear (e.g. use of the same data collection tools)
- Randomisation loses credibility if the pilot becomes national policy anyway (i.e. irrespective of results)

Implications for implementation



- Limited number of sites and projects
 - Creates smaller, i.e. more localised effects and tests fewer approaches (\neq early implementers)
- Scale of data collection
 - May affect the capacity in sites to develop their own ways of service delivery and diverts resources from service delivery towards the experiment (\neq pathfinder)
- The need to control variables (e.g. participant recruitment)
 - May result in reduced opportunities for local adjustment during the pilot; undermines local learning and ‘buy in’
 - May result in testing approaches that are to some extent an artefact of the pilot (i.e. Threat to external validity)
- Assumption of equipoise (genuine uncertainty about effects)
 - Is likely to diminish ‘buy in’ from pilot sites (\neq demonstrator)
 - Is at odds with the idea of sustainability (\neq early implementer)
 - Is impractical if there already is a commitment to continue (e.g. WSD, IB)

Further observations from the case studies

- There was a tendency to move from evaluating larger policy ambitions to specific interventions (e.g. 'whole system change', 'system transformation')
 - The question is whether 'whole system change' requires a different approach to evaluation
- Cost and outcome findings seem to have attracted much more policy interest than process findings
 - Use of other 'learning' much more patchy, but nonetheless present (e.g. incompatibility of income streams, addressing needs of older users in IB pilot evaluation)
 - However, this learning is relevant for implementation and it is relevant to sites (and non-sites)
- Piloting is best seen as contributing to longer term policy trends rather than representing a beginning and end point in the life of a policy

Questions for discussion



- Is it possible to determine the purpose(s) of piloting more clearly in the beginning? Can trade-offs be identified? Can purposes be held constant?
- What can/should we do better/differently in future vis-à-vis policy pilots and their evaluations?
- How do these findings relate to other pilots? Do they display similar tensions between purposes?