The politics and practice of policy piloting: three examples from health and social care in England

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

SSRG Annual Workshop, London, 27 March 2014

Outline

- Rationale for policy piloting and its critique
- The study
- The purposes of piloting
  - 'Typology'
  - Three pilot programmes
- Explaining the existence of multiple purposes
  - "What works" vs. piloting and evaluation as a strategy for persuasion

The 'official' rationale for piloting

- Improving policy and policy outcomes by using evidence to inform decision-making (EBPM)

- 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)

# Three observations indicate this may be problematic

- Complex interventions are difficult to evaluate
  - Findings from evaluation are multiple and not always easy to 'apply' to policy
  - Internal vs. external validity trade-off of RCTs (Cartwright and Hardie, 2012)

- Policy announced to be rolled out before pilots and evaluation have run their course

- Little evidence that findings from evaluations of policy pilots are (systematically) used to inform policy formulation

Limited capacity of governments to learn from failure

"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)

The three activities involved in piloting
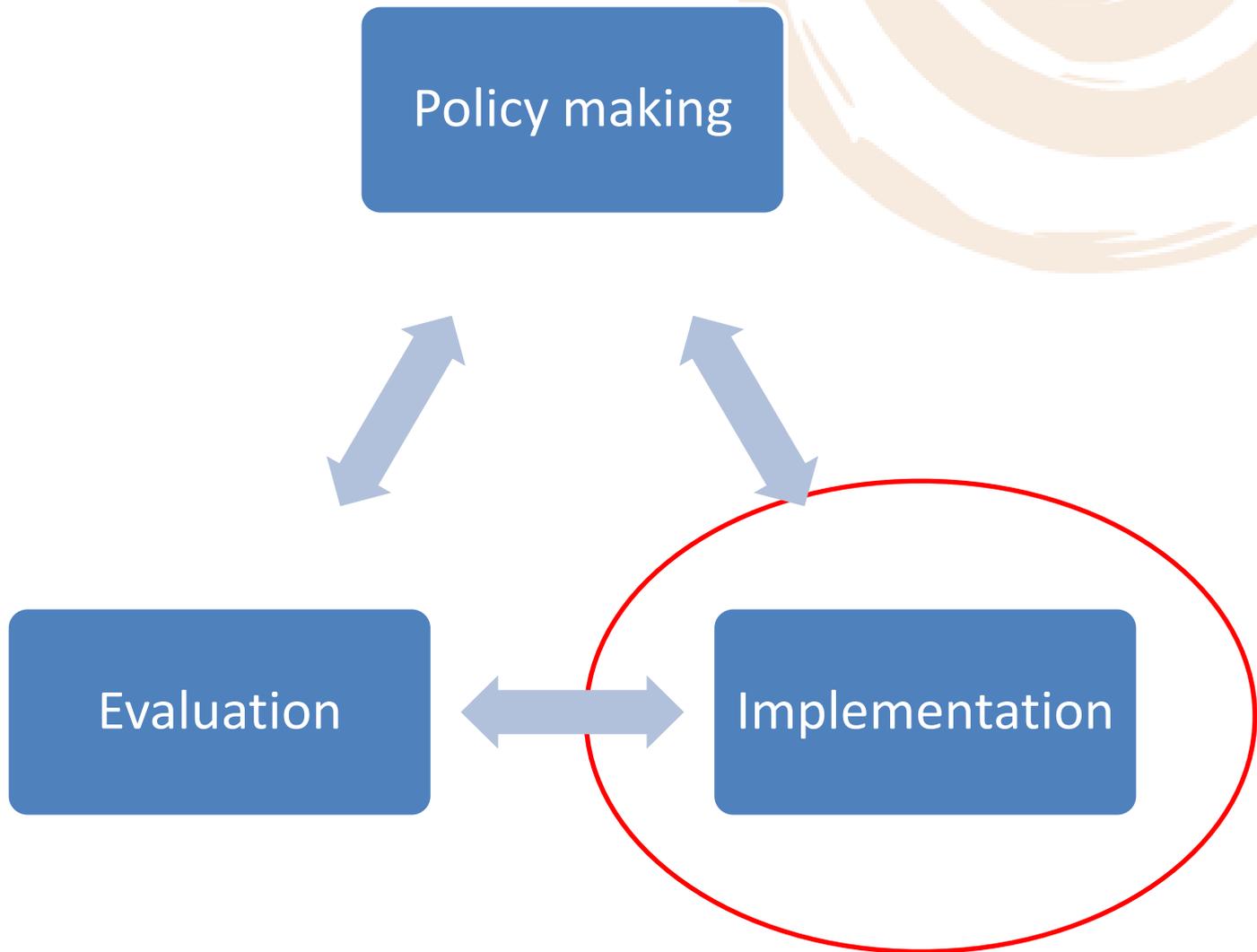
Aim and objectives

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

- Analysis of three case studies
  - High-profile policies (at the time); emerging from 2006 White Paper; pre-austerity funding; comprehensive evaluations commissioned

- Documentary analysis
  - 50+ documents, including policy documents, evaluation reports, media articles

- Interviews with key stakeholders
  - DH officials, evaluators, managers in pilot sites (n=31)

The three case studies
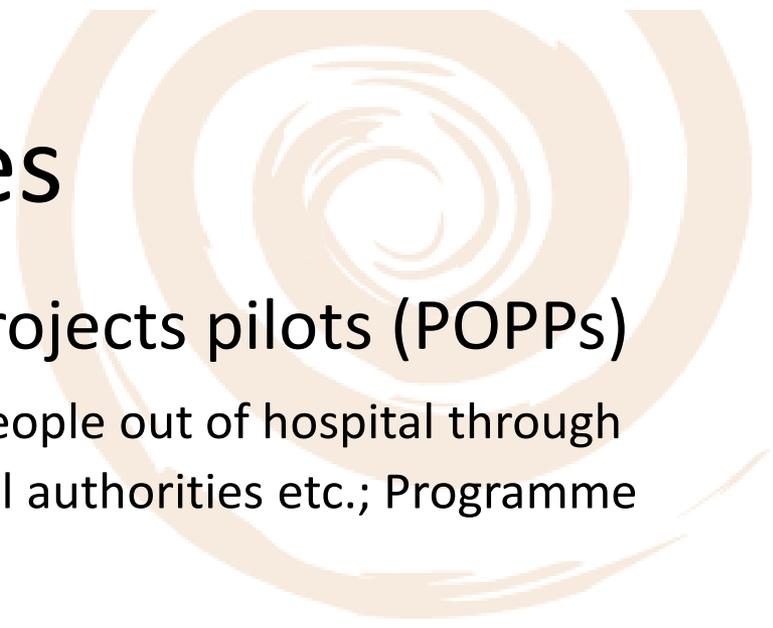
- Partnership for Older People Projects pilots (POPPs)
  - Interventions aimed at keeping older people out of hospital through working in partnership across NHS, local authorities etc.; Programme budget £60m.; 2006-09

- Individual Budgets 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

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 how the policy can be implemented successfully

- 'Pathfinder'/'Trailblazer'

  – Learning how to operationalise policy, how to overcome implementation barriers and how to improve processes and outcomes
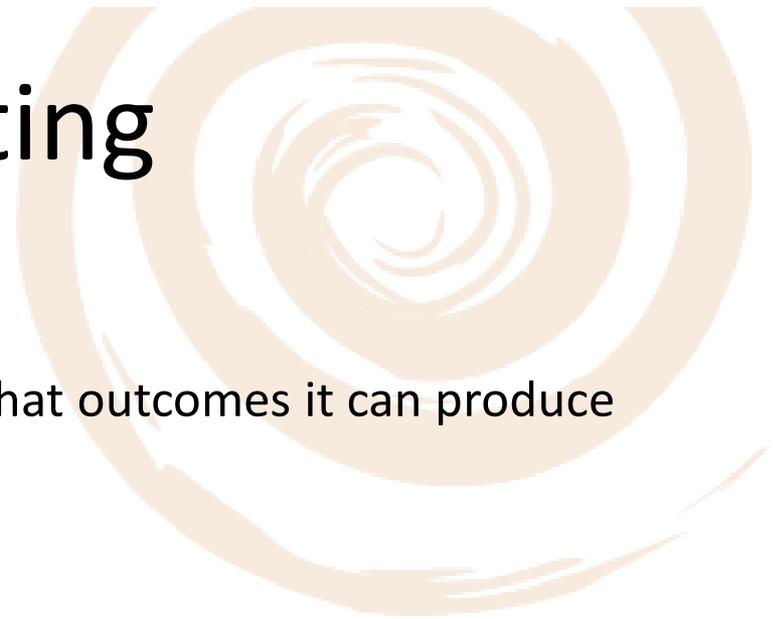
Partnerships for Older People Projects (POPP) pilot

- 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.
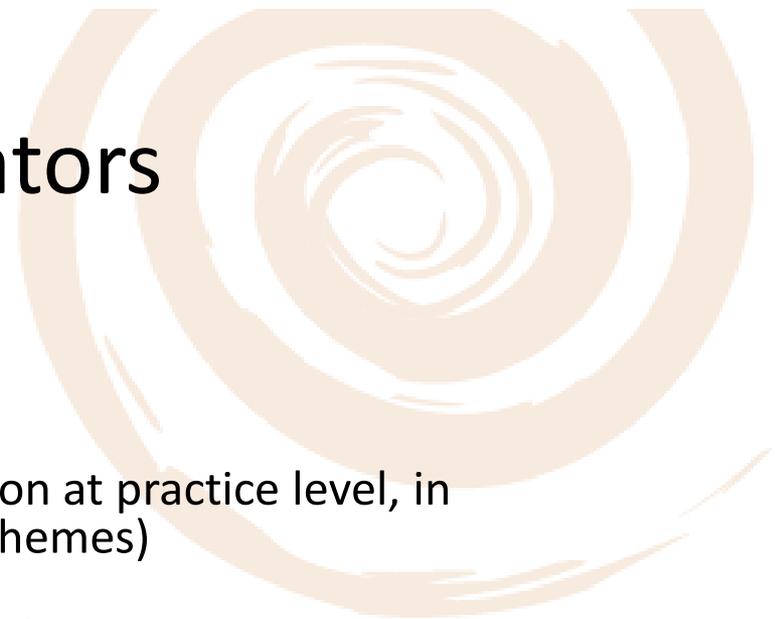
Individual Budgets pilots

- 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"
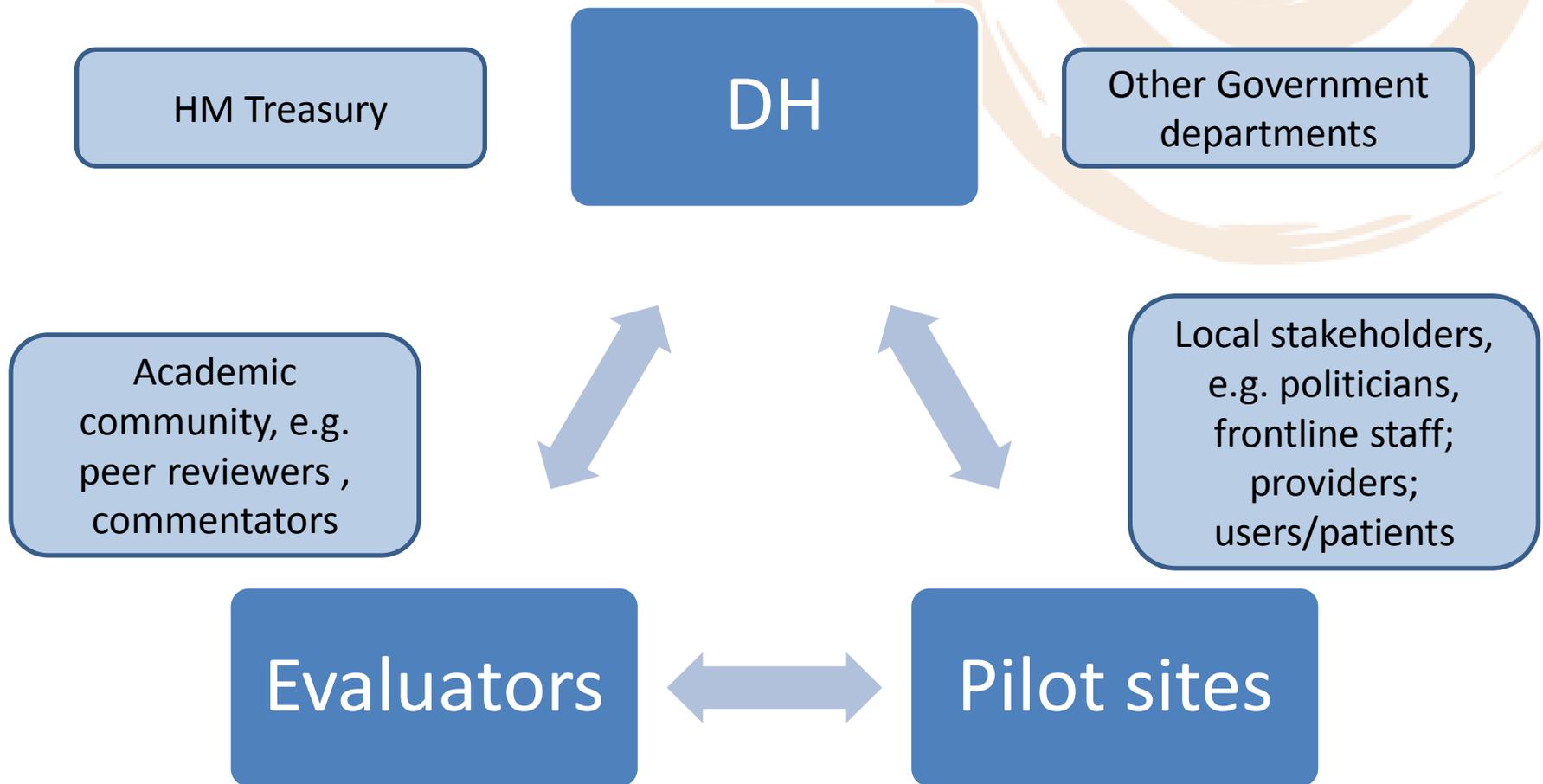
Whole System Demonstrators

- Was designed as an experiment

  - Sites (n=3), involving 6000 participants

  - '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

  - RCT said to be useful to persuade critical audience of the value of technologies

Stakeholder perspectives

HM Treasury

DH

Other Government departments

Academic community, e.g. peer reviewers , commentators

Local stakeholders, e.g. politicians, frontline staff; providers; users/patients

Evaluators

Pilot sites

Explaining multiple purposes

- Stakeholders in piloting had different objectives
  - Pilot sites: e.g. facilitating local change, access to funding, kudos
  - Evaluators: e.g. generating evidence
  - DH: e.g. input to policy development
- Purpose was not constant and motivation to pilot and evaluate changed over time ('goal drift')
- Lack of clarity of purpose at design stage

Explaining multiple purposes

Piloting for experimentation

- – Informing policy formulation
- – Preference for outcome research ("what works") and experimental research designs

vs.

Piloting for implementation and learning

- – Informing policy implementation
- – Unclear role of implementation research; process evaluation; formative evaluation

"Genuine uncertainty"?

- Policy commitments preceded the pilots
  - Targets to increase use of telecare in NHS Plan 2000
  - Long-term commitment to integration and to improving wellbeing of older people (e.g. PSA Targets)
  - Individual Budgets roll-out announced during evaluation

- Using the pilots to support the 'business case'
  - Evidence (effectiveness and cost) needed to be able to make the case for investment to the Treasury
  - Perception that "gold standard" of evidence of effectiveness (RCTs) could end debates "once and for all"

Proving the 'business case'
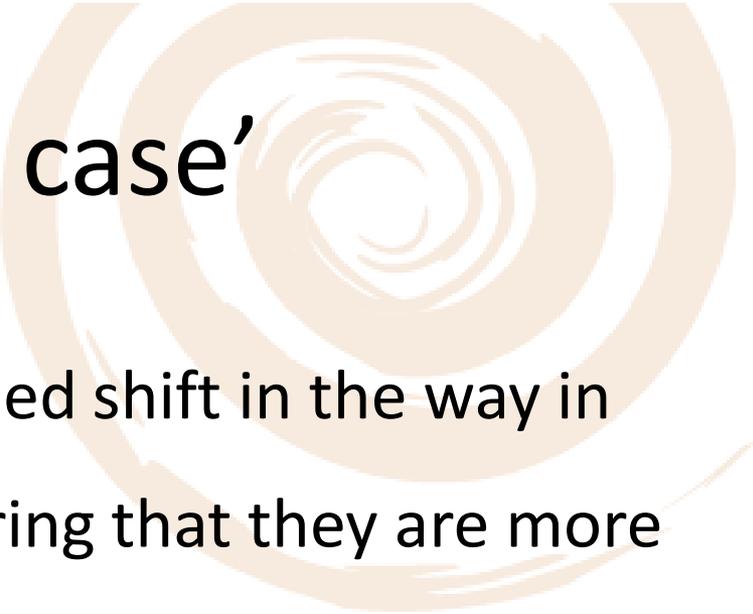
"There will be a radical and sustained shift in the way in which services are delivered, ensuring that they are more personalised and that they fit into people's busy lives. [...] The White Paper [Our health, our care, our say] proposed a number of whole system demonstrators that would **help us prove the business case for such wide ranging changes**." *[DH, Requirements, 2006: 5]*

Proving the 'case'

"Operational from 1 May 2006, the POPPs will **provide examples** of how innovative partnership arrangements can lead to improved outcomes for older people, particularly with respect to reduced hospital admissions and residential care stays. […] [new para] The economic case for primary and secondary disease prevention has been made. The task is now to develop local services that **translate this evidence into service delivery.** "

[DH, White Paper, 2006: 48]
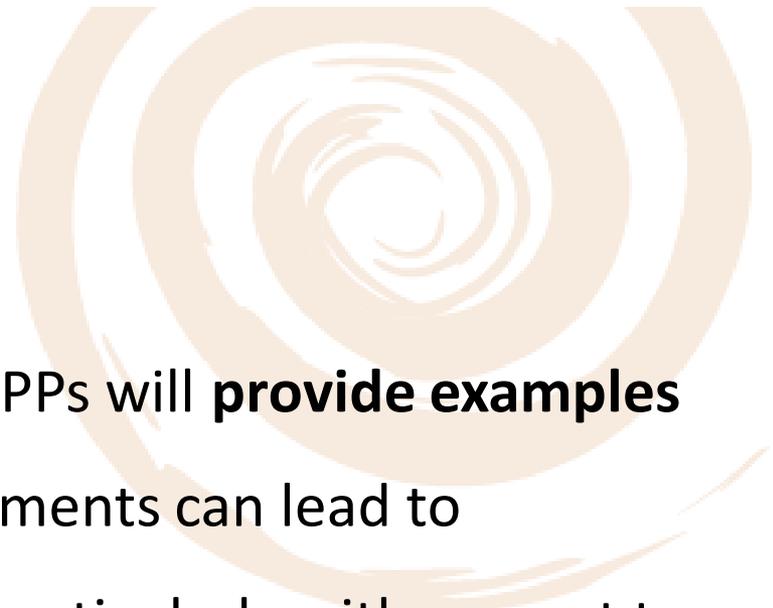
But what if … the findings are not supportive of the policy?

- Strength of the studies did not protect findings from 'politics' or preclude continued debate

- DH has a (limited) role in publication of findings set out in commissioning contracts (e.g. 28 day rule)

- …but there were examples of subtle influences on how findings were presented (e.g. Changing order of chapters; negotiating content of executive summary; publishing a response; adding own analysis to press communication

Health monitors to be installed in millions of homes

During the world's largest telehealth study, carried out in Newham in East London, Kent and Cornwall, about 120 lives were saved as a result of the technologies. Local trusts also spent 8 per cent less on each patient.

The NHS expects to spent £750 million on installing the systems, but says that it will save about £1.2 billion as a result over the next five years.

Mr Cameron said in a speech on medical innovation yesterday: "We've trialled it, it's been a huge success, and now we're on a drive to roll this out nationwide. This is going to make an extraordinary difference to people —
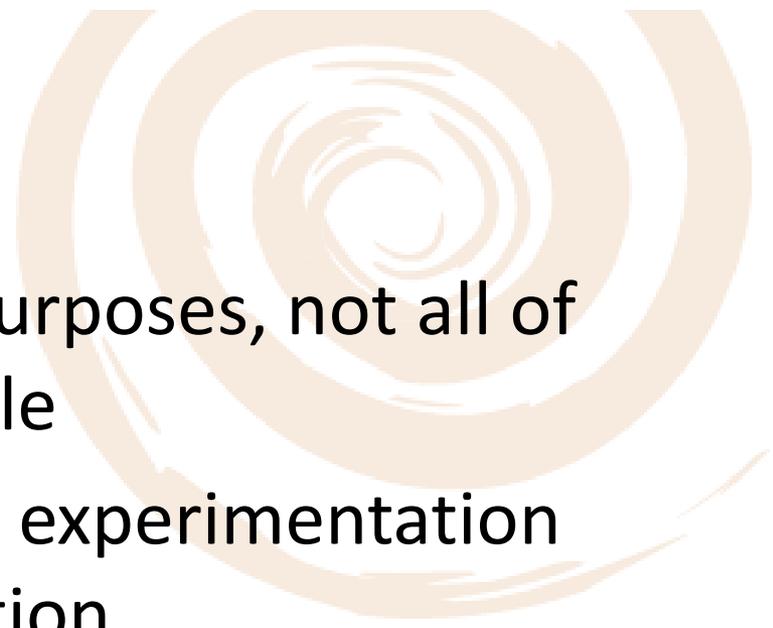
Journalist or DH based on Steventon et al. 2012 (bmj)

Steventon et al. 2012 (but not statistically significant)

DH analysts

Government interpretation

Conclusions

- Piloting can serve multiple purposes, not all of them are mutually compatible

- Tension between piloting for experimentation and piloting for implementation
  - RCTs as vehicle for persuasion rather than "genuine" test of whether policy "works"
  - Opportunities for local learning? Who learns what from national evaluation?
  - Problem of dealing with the findings from evaluation

- Useful for legitimating policy process, but limited impact on policy (and practice)

Thank you!

Disclaimer