PIRU POLICY INNOVATION RESEARCH UNIT

Policy experiments: Trade-off between rigour and relevance?

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

Social Research Association, 9 December 2013

Outline

- Definition(s)
- Study on policy piloting
- Three issues of RCT based policy experiments
  - Real life settings
  - Policy as intervention
  - Equipoise / uncertainty

What is a ‘policy experiment’?

Experiment:

“A study in which an intervention is deliberately introduced to observe its effects”

Shadish, Cook and Campbell, 2002: 12

Policy experiment:

a) Any form of testing and piloting of policy, with or without (formal) evaluation

b) Policy evaluation using experimental methods (e.g. randomised controlled trials)

The promise of the RCT

- RCTs focus on measuring outcomes and on establishing evidence of effectiveness; strong claims to validity of causal inference
- The expectations are that such findings
  - Are more relevant to policy and policy-makers (e.g. Haynes et al. 2012; Stoker, 2010), thus improve 'evidence use'
  - Produce more definite answers than non-randomised evaluations
  - Are less vulnerable to political contestation due to the strength of the evidence ('gold standard')
  - Give policy decisions credibility and added legitimacy

The study

- Aims
  - To examine the role of piloting in national policy-making in health and social care
- Design and methods
  - A qualitative study of case studies of DH policy piloting relating to health and social care
  - Selected to represent influential cases, i.e. affecting policy and academic debate
  - Documentary analysis (n=55); interviews with policy-makers, pilot site managers and evaluators (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 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

Key findings of the study

- Piloting was associated with multiple purposes, of which generating evidence was only one
- Other purposes were about implementation:
  - Early implementation, 'pump priming' (*pioneers*)
  - Demonstration to motivate others (*demonstrators*)
  - Fostering innovation locally
  - Operationalising policy (i.e. working out "the model"; *trailblazers; pathfinders*)
- There were tensions between different purposes!

Doing RCT-based policy experiments

1. Takes place in a ‘real life’ setting – will the implementation conditions be amenable to experimentation?
2. Needs a defined intervention – will the policy ‘mechanism’ be specific enough to allow inference of causality?
3. Requires equipoise – will there be ‘genuine uncertainty’ about policy effectiveness?

1. Real life setting: A ‘pragmatic’ trial

“So in that sense if you think pragmatic is just observing the world as it is and what would have happened, it isn’t. As soon as you introduce the concept of randomisation and control it is complete anathema to the behaviour in health and social services so that is an element which isn’t pragmatic, but elements of the trial were pragmatic. So selection wasn’t pragmatic because we want to be clear who will be selected, randomisation was not pragmatic at all, but the way they supported services through Telehealth and Telecare and the models they used and indeed the devices they used, although there were set criteria, they defined them.”

Researcher

The perspective of implementers

- Sites participate on a voluntary basis and received resources from DH for this purposes
- Their motivation to participate was to effect local change (and to be seen doing so)
- Issues were mostly about control groups and randomisation (e.g. eligibility criteria different from mainstream service; existing users)
- Felt they lacked flexibility to adjust to experience made during the trial

2. Working out ‘the model’

- Policy is different from other interventions; “more gradual perturbations over long periods in a complex system” (Hawe, Shiell and Riley, 2009)
- There was no ‘model’ or ‘one’ intervention
  - IB: operating model in the beginning had to be created, by discarding other options
  - POPP: multitude of different projects (146 ‘core’, over 500 other projects)

3. ‘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 RCT to support the ‘business case’
  - Evidence (effectiveness and cost) needed to be able to make the case for investment to the Treasury
  - Perception that robust evidence of effectiveness 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]

Making the case – persuading critical audiences

"How do you get your medical world, the GPs, the nurses, all of those, how do you get people who are currently working one way to work differently? [...] What is going to persuade them? Proper evaluated evidence."

Official

But what if the findings are not supportive of the policy?

- Strength of the studies did not protect findings from ‘politics’ or preclude political 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

Rigour and relevance – a trade-off?

- A false dichotomy – both determined by perceptions of policy-makers of what supports an argument
- Case studies suggest that relevance of the experiments was in supporting a case for policy rather than to question it
- They also suggest that emphasis of rigour in terms of doing RCTs introduced new risks (implementation; reputational risks)

Thank you!

Disclaimer

The Policy Research Unit for Policy Innovation Research is funded by the Department of Health. The views expressed in this presentation are those of the researchers alone and do not necessarily represent those of the Department of Health.