

# Seven challenges to developing rigorous policy evidence

Mark Petticrew

Faculty of Public Health and Policy

London School of Hygiene and Tropical Medicine (LSHTM)

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



# **Seven challenges to developing rigorous policy evidence**

- 1. (Lack of) Timeliness**
- 2. An epidemic of certainty (Absence of “equipoise”)**
- 3. Narrow views of what “evidence-based” means**
- 4. “Evalu-mania”: Unrealistic demands from researchers to evaluate everything that moves**
- 5. The odds are stacked against evidence use**
- 6. Communication difficulties between researchers and policy users**
- 7. The laws of evidence production**

# 1. (Lack of) Timeliness



- Policy “windows”
- But...surely important policy issues are constants (what should we do with criminals; does welfare-to-work “work”; do early years interventions work). Not just one opportunity to be relevant.
- Policy “windows” reappear regularly
- Good evidence has a long shelf life: U.S. Head Start evidence used to drive interventions in the UK and the other countries

## **(Lack of) Timeliness; sometimes action can't wait for the full academic process**

Colin Blakemore (The Times, October, 2010):

*“What do we want?”*

*“Thorough consideration of the evidence that public expenditure on research is causally linked to economic benefits!”*

*“When do we want it?”*

*“As soon as the Government is able to gather full, peer-reviewed data!”*

## Buxton's Law

- Getting the timing right is difficult:
- “It is always too early [for rigorous evaluation] until suddenly it's too late.” [Buxton, 1987]

- 1: “Good evidence may sometimes be expensive and slow but it is an investment; research done well has a long shelf life”
- 2: Produce “quick and clean” evidence: Rapid reviews, rapid evidence assessments...
- 3. Interim outputs and continual engagement

## 2. An epidemic of certainty (or “Absence of equipoise”)

- “Research requires uncertainty (Equipoise):

*“[clinical equipoise exists when] there is no consensus within the expert clinical community about the comparative merits of the alternatives to be tested”  
(Freedman, 1987)*

- Once uncertainty about the correct course of action has been acknowledged, it is easier to admit the need to do a trial to identify effective alternatives
- The concept of equipoise is not much discussed outside of clinical trials. It may be useful in non-clinical areas *and it might also apply to non-trial research*
- Instead, it often looks like there is an epidemic of political certainty; uncertainty about policy alternatives is rarely acknowledged which makes it difficult to do RCTs. In fact, policymakers are experimenting on us all the time

# Uncontrolled experimentation is the norm

In fact, all new initiatives (like new policies) are potential experiments – and the public are frequently ‘enrolled’ in these real-life policy ‘**experiments**’ without any real prospect of anyone learning anything substantial about the effects of those interventions

The former Health Secretary **Kenneth Clark** famously called the **NHS** ‘a great **experiment**’, and he also referred to his health service reforms in the early 1990s in these same terms.

Nigel Lawson, also described his macroeconomic policies as ‘The **British Experiment**.’





## REFORMS AS EXPERIMENTS<sup>1</sup>

DONALD T. CAMPBELL<sup>2</sup>  
*Northwestern University*

The United States and other modern nations should be ready for an experimental approach to social reform, an approach in which we try out new programs designed to cure specific social problems, in which we learn whether or not these programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available. Our readiness for this stage is indicated by the inclusion of specific provisions for program evaluation in the first wave of the "Great Society" legislation, and by the current congressional proposals for establishing "social indicators" and socially relevant "data banks". So long have we had good intentions in this regard that many may feel we are already at this stage, that we already are continuing or discontinuing programs on the basis of assessed effectiveness. It is a theme of this article that this is not at all so, that most ameliorative programs end up with no interpretable evaluation (Etzioni, 1968; Hyman & Wright, 1967; Schwartz, 1961). We must look hard at the sources of this condition, and design ways of overcoming the difficulties. This article is a preliminary effort in this regard.

Many of the difficulties lie in the intransigencies of the research setting and in the presence of recurrent seductive pitfalls of interpretation. The bulk of this article will be devoted to these problems. But the few available solutions turn out to depend upon correct administrative decisions in the initiation and execution of the program. These decisions are made in a political arena, and involve political jeopardies that are often sufficient to explain the lack of hard-headed evaluation of effects. Removing reform administrators from the political spotlight seems both highly unlikely, and undesirable even if it were possible. What is instead essential is that the social scientist research advisor understand the political realities of the situation, and that he aid by helping create a public demand for hard-headed evaluation, by contributing to those political interventions that reduce the liability of honest evaluation, and by educating future administrators to the problems and possibilities.

<sup>1</sup> The preparation of this paper has been supported by National Science Foundation Grant GS1309X. Versions of this paper have been presented as the Northwestern University Alumni Fund Lecture, January 24, 1968; to the Social Psychology Section of the British Psychological Society at Oxford, September 20, 1968; to the International Conference on Social Psychology at Prague, October 7, 1968 (under a different title); and to several other groups.

<sup>2</sup> Requests for reprints should be sent to Donald T. Campbell, Department of Psychology, Northwestern University, Evanston, Illinois 60201.

**Admitting  
uncertainty about  
evidence can be  
enormously  
politically helpful to  
decisionmakers**

Campbell, DT., *American Psychologist* (1969) 24:409-429

**Campbell D, Reforms as Experiments**  
***American Psychologist* (1969) 24:409-429**

“It is one of the most characteristic aspects of the present situation that specific reforms are advocated as though they were certain to be successful...

If the political and administrative system has committed itself in advance to the [...] efficacy of its reforms, it cannot tolerate learning of failure...We must be able to advocate without that excess of commitment that blinds us to reality testing...

...One simple shift in political posture which would reduce the problem is the shift from the advocacy of a specific reform to the advocacy of the seriousness of the problem, [and] advocacy of persistence in alternative reform efforts should the first one fail. The political stance would become: ***“This is a serious problem we propose to initiate Policy A on an experimental basis. If after 5 years there has been no significant improvement, we will shift to Policy B”...***

Negative results, a failure of the first programme, would not jeopardize his job, for his job would be to keep after the problem until something was found that worked.

### **3. Narrow views of evidence-based policy**

- Evidence “based”...
- Evidence “informed”...
- Evidence “aware”...

- Evidence “based”...
- Evidence “informed”...
- Evidence “aware”...
  
- *Subliminal use of evidence (“I’ve a sort of a feeling I saw some evidence on this, but can’t be sure”)*
  
- *Homeopathic use of evidence (Where the presence of evidence in a policy can not be detected by normal processes)*

## **Evidence-based decisionmaking: researchers' narrow views of what this means are also a barrier**

- It is not just about (1) RCTs, and(2) not solely about research evidence

- *“There is growing concern that a standardized approach based on what is, in effect, a narrow definition of science is compromising the use of diverse research designs and multidisciplinary methodologies needed to respond effectively to solving “context-specific” public health problems”*

*(Behague et al., Soc Sci Med 2009)*

## Standardisation

- There is still a widespread view that evidence-based decisionmaking formalises, standardises and limits interventions, reducing the role of other forms of information (including values), and prevents flexibility in implementation
- A bit like “Cookbook medicine”:
- *“Evidence based medicine is not “cookbook” medicine. Because it requires a bottom-up approach **that integrates the best external evidence with individual clinical expertise and patients' choice**, it cannot result in slavish, cookbook approaches to individual patient care”.*

Evidence based medicine: what it is and what it isn't. (Sackett et al. BMJ, 1996)



- “Evidence based medicine is not restricted to randomised trials and meta-analyses. It involves tracking down the best external evidence with which to answer our clinical questions... when asking questions about therapy that we should try to avoid the non-experimental approaches, since these routinely lead to false positive conclusions about efficacy”.

- “Because the randomised trial, and especially the systematic review of several randomised trials, is so much more likely to inform us and so much less likely to mislead us, it has become the “gold standard” for judging whether a treatment does more good than harm. However, some questions about therapy do not require randomised trials ...or cannot wait for the trials to be conducted. And if no randomised trial has been carried out for our patient's predicament, we must follow the trail to the next best external evidence and work from there”.

**Evidence based medicine: what it is and what it isn't. (Sackett et al. BMJ, 1996)**

# So...

- EBP is not solely about research, or indeed RCTs
- EBP does involve prioritising the use of best available (RCTs where these are available, or when new ones can be done) to answer questions about the effectiveness of interventions
- Other evidence needs to be used to address other types of question (questions about prevalence requires surveys; questions about acceptability, or understandings often require rigorous qualitative research; and so on)

#### **4. “Evalu-mania”: Unrealistic demands to evaluate everything that moves. Do we really need to evaluate everything?**

- “Few may disagree in principle with the evaluative ‘call to arms’, but its implementation raises a number of scientific, practical and prioritisation issues” (*Ogilvie et al. 2011*)
- Richard Windsor has noted there is a need to ensure that an evaluation is ‘realistic, prudent and efficient’ because ‘Every component of a program usually cannot and, in most cases, should not be evaluated’, emphasising that decisions need to be made about how best to deploy scarce resources for evaluation in order to maximise the new learning and evidence produced (*Windsor, 2004*).

## Some arguments why we may not want to evaluate everything...

- Limitations in terms of funds, and researcher capacity
- Policy-makers' may not use the findings - the main facilitators of research evidence in policymaking are personal contact, timeliness, clear recommendations, good quality research, confirmatory research and community pressure for research. By contrast, research that included effectiveness data was mentioned in only 3 studies out of 24 (*Innvaer et al. 2002*)
- The “Value of Information” from all intervention studies is not the same (some may have little impact on a real decision)
- A “gap” is not the same as a “need” – in some areas (**which?**) we may have already enough evidence

## 5. The odds are stacked against evidence being used...

- No guarantee that the findings will be useful, or used: At least 293 barriers to using evidence have been identified (Cabana, 1999)
- One of these barriers is that evidence can be politically unacceptable or at least unhelpful:

## Politically unhelpful evaluation

- “... there are costs to having really good evaluations not just the financial cost, they do cost more obviously but you know, if ... it’s really good and the results you know tell you that your intervention isn’t working then you’re in trouble, and I think to some extent ...people would rather have you know vaguer information about processes, which ...carries less risk of being hostages to fortune to some extent...”

**Researchers do not give policymakers clear advice on when to use particular research methods, and experiments are often downplayed:**

*“The message you get from the research community is very ambivalent...[there] is a very strong strand of thinking ... which you know tends to be suspicious of using experimental methods in relation to social interventions... that influences the thinking of [...] research brokers within government. ... Policy makers are getting these kind of rather muffled messages about when to do a trial or ... when some other method will do... so it’s hardly surprising that they are you know, quite happy to ... go on using weaker methods, ...they’re not getting a clear steer.. the choices they’re given don’t involve the option of running a trial. ...There are... quite influential papers about the methodology in evaluating social interventions which give policy makers a lot of rope to hang themselves”.*



- “Certainly in British politics, the power of a story beats almost anything.” (Policy advisor, UK)
- All this suggests that the ethical and methodological barriers are only part of the story
- Descriptive evaluations are also politically more expedient/more acceptable

## **6. Communication difficulties between researchers and policy users**

## 7. Laws governing evidence production

- **The inverse evidence law: the strongest evidence we have is often about risk factor modification, and we have more, “weaker” evidence about many of the wider social economic and environmental determinants of health (including policies)**
- E.g. The evidence on modal shift in transport - how to get people to walk and cycle more...

## Types of intervention

- “Health promotion” activities (*Education campaigns; free bikes*)
- Engineering measures (*Bicycle infrastructure; traffic restraint*)
- Financial incentives (*voucher/fine to leave car at home*)
- Providing alternative services (*e.g. A new railway station*)
- Complex urban transport policies

*“...some interventions have only been evaluated in non-randomised studies”*

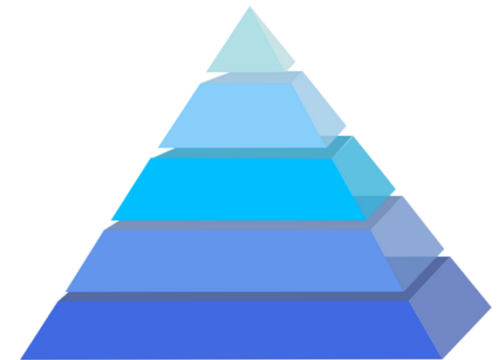
## Transport policies and physical activity

Increasing  
scale/complexity ↑

- Complex urban transport policies
- Providing alternative services (*new railway station*)
- Engineering measures (*Bike infrastructure; traffic restraint*)
- Financial incentives
- “Health promotion” activities (*Education campaigns; free bikes; leaflets*)

# Study designs used to evaluate interventions to bring about modal shift (Ogilvie et al., 2007)

	N (studies)
Randomised controlled trial ( <i>individual-level</i> )	3
Panel survey	13
Repeated cross-sectional survey (community-level)	17
Retrospective or after-only survey	11
Case study /uncertain ( <i>city-level</i> )	20



**Inverse evidence law:** the strongest evidence we have is often about risk factor modification, and we have more, “weaker” evidence about many of the wider social economic and environmental determinants of health (including policies)

## **And finally: What does social science contribute to the development of rigorous policy evidence?**

- **Methods:** The methodological tools of evidence-based policy derive from the social sciences, including RCTs, and systematic reviews
- **A social science perspective:** evidence production and use are social processes. We need to understand these if we are to produce “evidence” that is of value to users, and understand and enhance its use