

What Is Evidence-Based Policy?*

Paul H. Jensen

**Melbourne Institute of Applied Economic and Social Research
The University of Melbourne**

Melbourne Institute Policy Brief No. 4/13

ISSN 2201-5477 (Print)

ISSN 2201-5485 (Online)

ISBN 978-0-7340-4321-4

August 2013

* This Policy Brief has benefited from comments from Deborah Cobb-Clark, Beth Webster, Gaétan de Rassenfosse and James Davis. Email <pjensen@unimelb.edu.au>.

**Melbourne Institute of Applied Economic and Social Research
The University of Melbourne**

Victoria 3010 Australia

***Telephone* (03) 8344 2100**

***Fax* (03) 8344 2111**

***Email* melb-inst@unimelb.edu.au**

***WWW Address* <http://www.melbourneinstitute.com>**

Melbourne Institute Policy Briefs Series

Commencing in 2013 the Policy Brief Series is a new collection of research publications that will be produced by the Melbourne Institute. This series will examine current policy issues and will provide an independent platform to examine pertinent issues in public debate.

Log onto melbourneinstitute.com or follow @BusEcoNews for more details on this and other series from Australia's leading and longest standing research institute in the field of economics and social policy.

Abstract

Around election time, politicians often call on the need for more ‘evidence-based policy’. This message is simple, but the importance of being able to distinguish strong evidence from weak evidence is often forgotten. In this Policy Brief, the rationale underpinning the ‘evidence-based’ approach to public policy is carefully explained, as are the pros and cons of the different methods used to construct the evidence base.

JEL classification: A11, D04

Keywords: Evidence, public policy, counterfactual

“You change your laws so fast...without inquiring after results past or present, that it is all experiment, seesaw, doctrinaire; a shuttlecock between battledores.”

Florence Nightingale, as quoted by Gary Banks in Productivity Commission (2010, p. 3), *Strengthening Evidence-based Policy in the Australian Federation*, Commonwealth of Australia, Canberra.

Introduction

The ultimate goal of evidence-based policymaking is better public policies, thereby creating healthier and wealthier societies. Moreover, evidence-based policies should provide taxpayers with more confidence that the Government is spending their hard earned money wisely. Once we agree that these are the right goals to strive for, the question is: how do we get there? Setting lofty, long-term goals is no doubt important, but there is much we can do in the short-term to evaluate how government policies are faring with regard to our ‘healthy and wealthy’ agenda.

One important way forward relates to evidence-based policy. The idea is simple: we don’t always know which government policies are going to work (since the world is complex and there are always lots of unintended consequences of our policies, as well as the intended ones), so let’s try and determine which ones ‘work’ and which ones don’t. The ultimate objective of ‘evidence-based’ policy is to use actual evidence on what works – rather than rely on ideology – to promote good public policies. That way, good policies survive and bad policies are killed off. Using an incremental approach like this, we can continue to make the world a better place.

Despite its apparent simplicity, the evidence-based policy agenda is often misunderstood. Of course, there are sometimes some disagreements about the outcomes we are interested in (i.e. what we mean when we say the policy ‘works’). Once these issues have been resolved, we need a set of tools and techniques to use in our determination of the effectiveness of the policy. The key challenge is to identify a counterfactual: what would have happened in the absence of participating in the program? Although this is challenging, there are ways in which a counterfactual can be calculated. And once these challenges are embraced, the process of designing and implementing better public policy can continue apace.

Building the Evidence Base

Public policymakers face a difficult task in trying to maximise the returns from their programs. If we abstract from the political issues which shape the existence or size of programs and simply focus on the economic (non-political) objectives of the program, it seems obvious that all policymakers want the best possible outcomes given the prevailing economic climate (which changes over time).

There are lots of problems in achieving the best possible outcomes: for example, there are often ‘unintended consequences’ resulting from specific government programs (e.g. there is evidence that parents changed the timing of births via inducements and caesarean sections as a result of the Federal Government’s baby bonus). More generally, in most instances, it just isn’t possible to predict the effects of a new policy, unless similar policies have been implemented elsewhere.

However, we can only learn from others' experiences if the objectives and outcomes are rigorously documented: the quality of the analysis undertaken on a program directly influences the quality of the lessons learned. Since it is extremely difficult to establish causal effects of policies, such analysis must be designed and performed very carefully. Abstracting from good luck, better evidence is a necessary but not sufficient condition for better public policies.

Doubtless, there are many factors that may influence the success of a specific policy (e.g. even 'good programs' could produce 'bad' results during a downturn in the business cycle). Nevertheless, there are a range of methods and techniques – from randomized controlled trials to difference-in-differences and case studies – that can be used to examine the effects of specific programs. But because not all evidence produced by these methods is the same quality, it is equally important to be able to rank-order them. That is, to provide a hierarchy of evidence – in which the gold standard will provide robust *causal* evidence of the effects of a specific program.

If evaluation is difficult – and individual Government departments have strong attachments to specific programs – why is it worthwhile improving the quality of the evaluations? One obvious reason is that in order for programs to be supported by the central agencies, it is important to provide them with evidence of sufficiently high quality that they believe the results. In the industry policy domain, for example, this is crucial since there is general scepticism regarding the need for government support of industry. If this isn't tackled directly, industry policy will most likely suffer in terms of future budget allocations, particularly as other policy areas – including education, health and social policy – are constantly moving closer to the best-practice evaluation frontier.

In recent years, it has become clearer that access to unit-record data has become important for building the evidence base: the analyst must be able to see the characteristics of the individual unit of interest (e.g. firm or household) in order to conduct rigorous analysis and draw strong inferences about the effects of policy. And with many new initiatives to link data and make it more accessible, there is every chance that technology will make it much cheaper (and safer in terms of privacy) to do so in the future. For example, there are exciting new initiatives such as the NORC platform at the University of Chicago which promote access to confidential data.

Hierarchies of Evidence

Almost all policies involve tradeoffs: there are winners and losers (costs and benefits) associated with policy implementation. Most policymakers adopt a utilitarian approach to understanding the effects of policy introduction: the metric applied to the policy is whether it results in 'the greatest good for the greatest number' (to paraphrase Jeremy Bentham).

Economy-wide analyses of the policy are of interest, but are extremely difficult to address and typically rely on controversial assumptions. Such analyses acknowledge that policies have effects that ripple throughout the economy, not just on the target group of interest. For example, programs designed to impact mothers' labour supply will probably affect mothers' decisions about if (and when) to return to (casual, part-time or full-time) work, but they are also likely to have effects on the demand for childcare services, take-away meals and dry cleaning services. In these studies, the determination of whether a program 'works' is in the broadest sense of the word since it captures all of the economic consequences of participation in the program. Other studies focus on the simpler

(but still difficult) task of evaluating the effects of the program on the target group (i.e. that the program works in the 'narrow sense'). What this approach attempts to do is to identify the *net causal effect*.

In attempting to construct a counterfactual, one of the key issues is that people generally choose whether to participate in a government program. This creates problems when trying to identify the effects of program participation as explained below. The following strategies could be used by the analyst:

- i) Analyse individuals before and after participation in the program (referred to as the 'treatment' group). However, this 'time-series' approach will not help the researcher disentangle the effects of the treatment from other factors (e.g. upswing in economic activity for some people due to global forces).
- ii) Observe two identical individuals/firms: one who receives government assistance (the treatment group) and one who doesn't receive government assistance (the control group) and observe them over time. However, it is impossible to have two identical individuals/firms. If there are systematic differences between the two, then it is difficult to disentangle the effects of the policy from differences in the individuals (which may be unobserved).
- iii) Observe the treatment and control groups before and after the participation in the program. But this doesn't enable the analyst to separate the effects of the program from other contemporaneous factors.
- iv) Observe the same individual in two different states of the world at the same time (which is impossible).
- v) Take advantage of a quasi-natural experiment where one group is subjected to an exogenous effect and an otherwise identical group is not. By 'exogenous' we mean the affected people had no choice in whether or not they are part of the treatment group and are not systematically different (in relation to possible confounding factors) from the control group.
- vi) Randomly allocate individuals to either a treatment or a control group and compare them over time. This is the gold standard.

With the exception of the impossible option iv), each of these strategies can be categorised as using either an experimental approach or a non-experimental approach. The disagreements between the two camps – often referred to as the 'randomistas' and the 'regressionistas' respectively – are quite marked. The strengths and weaknesses of both approaches are discussed below.

Regressionistas vs Randomistas

The experimental approach in social science (e.g. randomised controlled trials) mimics the approach adopted in the natural sciences by randomly assigning individuals/firms to a 'treatment'. By flipping a coin to determine whether an individual is in the treatment or control group, it is possible to overcome the problems associated with systematic but unobserved differences (e.g. differences in how motivated the individuals are) in the 'treatment' and 'control' groups. This relies on the fact that the sample size is large enough, since the estimate of the effect will converge on the true mean as the sample size increases.

The question that remains is what limitations there are with regard to the ability of the experimental approach in social science. One obvious pitfall in the social sciences is the lack of a 'placebo': in medical trials, two groups are given pills, but one is given a pill which turns out to have no active ingredient. This approach simply can't be imitated in the social sciences: it is impossible to fool the members of the control group into thinking they might be receiving a treatment when they aren't.

By contrast, the non-experimental approach uses econometrics to try and deal with the self-selection problem. Selection into the program on factors that we cannot observe directly (which may be correlated with variables of interest) is indeed a major problem. But with the advent of more (and cheaper) data, this can be effectively handled by econometric techniques. If so, the marginal benefit of experiments over non-experimental methods should be diminishing over time.

In order to shed light on the pros and cons of the two approaches, the following commonly-reported critiques are stated and evaluated.

Ethical issues. The counter to the proposition that it is unethical to simply toss a coin to determine who receives the 'treatment' is that the reason the randomised trial is taking place is to determine whether the policy works: it is only unethical to conduct the trial *if we already know that the policy works*. If you don't know whether a specific policy works, it is unethical i) to do nothing; or ii) not to conduct an experiment to find out. Moreover, in the case of a specific program that is over-subscribed (i.e. there are more people who want to participate than there are spaces available), then the ethical issues seem much less acute: since someone has to miss out, it may as well be determined on a random basis (that is as fair as can be achieved).

However, there is residual concern that experiments have been used in areas where we do know whether the policy works: for example, de-worming tablets. On top of this, there are issue about 'informed consent' since many people in aren't asked whether they would like to be part of an experiment.

Practical issues. Experiments can't be used in every context. For example, it is impossible to design and conduct an experiment on macroeconomic issues such as a random shock to interest rates. Of course, such an experiment could be designed and implemented in a laboratory setting, but that is not the focus of most experiments. In addition, a fascination with experiments may lead researchers to avoid important policy issues that can't be solved using experiments. That is, researchers might seek good answers instead of good questions.

Generalisability issues. Experiments are typically conducted in environments with unique characteristics which may not be representative of all possible environments. Therefore, the results observed in one setting might not be generalizable to all contexts. Problems of this nature arise in non-experimental analysis too. But experiments tend to get criticised for this shortcoming more than other methodological approaches simply because experiments have solved most of the other methodological issues! The question is whether the external validity issues are greater in experiments than in non-experiments.

With regard to 'internal validity' – that is, whether you are estimating the causal effect of interest – experiments seem to outperform non-experiments. The correct weight to be applied to internal validity versus external validity (assuming there is some tradeoff between the two) is unclear: many studies tend to favour striving for greater internal validity, but it is unclear at what cost this comes. However, it is clear that in economic analyses, the issues of external validity are much more acute

than say in biomedical research. In other words, a bioactive agent is likely to work in Africa if previously shown to work in England. The same is not true of most economic policy intervention since the culture, institutions and norms in the two environments are quite different.

Statistical issues: there are two stages to the process of determining the ‘treatment’ and ‘control’ groups. Take a population of units (individuals/firms) from which you want to draw the two groups. The 1st stage is to select the ‘treatment panel’: those individual units which are willing to be part of the experiment. The 2nd stage involves randomly allocating each of the units in the treatment panel to either the ‘treatment’ or the ‘control’ group. One of the virtues put forward by advocates of experiments relates to the fact that they are free of (self-) selection bias. But this is only true with regard to the 2nd stage of the process noted above: in the 1st stage, it is necessary to select which units in the population will participate in the experiment and this *might not* be done randomly.

Also note that experiments provide an average effect (not a median effect, and not a percentage of people whose position improved). So, just because a policy produces a positive effect on average doesn’t mean that everyone participating in the program will experience the average effect. Of course, there is a distribution around the average: and if the distribution is spread widely (i.e. there is a high variance), the performance of a given individual could be much better (or much worse) than the average. Indeed, a result which showed that there was an average positive effect of a program could be dominated by a few winners (who win big) and many who fair much worse.

Substitution issues. One final issue relates to the behaviour of the members of the control group: in some situations, it is possible that they will seek out alternative substitutes to the treatment (since, as we noted above, one of the weaknesses of experiments in social science is that there is no placebo given to the control group). That is, if they believe that they have been ‘denied’ a potentially-valuable treatment, they will seek out an alternative. This potentially dilutes the experiment since the control group has now modified its behaviour from the desired neutral set-up intended by the experiment— it has been ‘pseudo-treated’.

Conclusions

Evaluating the nation’s economic and social policies is important if we are to continue to live in a prosperous nation. Moreover, ensuring that we spend taxpayers’ money wisely and prudently is an important part of the covenant between government and the people. Once we accept that evaluation is important, the question is: how should we do it? In this Policy Brief, the state-of-the-art as it relates to evaluation methods – and the rationale underpinning them – has been articulated. The simple answer to the question posed is that there is no silver bullet that can be applied in all contexts. Although there are lots of good reasons to support randomised controlled trials, there are also lots of important limitations to their use in social contexts. That is, there is a range of interesting policy questions that randomised controlled trials probably cannot answer. These issues should not be overlooked when designing and implementing innovation policy program evaluations.

Nevertheless, there is more that could be done to improve the quality of the evidence base used in designing public policy in Australia. For example, we should renew efforts to work with the Australian Bureau of Statistics to obtain access to unit-record data, think carefully about how to design and implement program evaluation, continue to build capability within the Government on policy evaluation, and promote the adoption of best-practice evaluation methodologies.

