



Alliance
for Useful
Evidence

THE CHALLENGES OF EVIDENCE

PROVOCATION PAPER FOR THE
ALLIANCE FOR USEFUL EVIDENCE,
NESTA

Dr Ruth Levitt

November 2013

This is a paper for discussion.

The author would welcome comments, which should be emailed to: **ruth.levitt@gmail.com** or **Jonathan.Breckon@nesta.org.uk**

The paper presents the views of the author and these do not necessarily reflect the views of the Alliance for Useful Evidence or its constituent partners.

The Alliance champions the use of evidence in social policy and practice. We are an open-access network of individuals from across government, universities, charities, business and local authorities in the UK and internationally. The Alliance provides a focal point for advancing the evidence agenda, developing a collective voice, whilst aiding collaboration and knowledge sharing, through debate and discussion. We are funded by the BIG Lottery Fund, the Economic and Social Research Council and Nesta. Membership is free. To sign up please visit: **www.alliance4usefulevidence.org**

Nesta...



THE CHALLENGES OF EVIDENCE

Which of these statements is true?

Evidence is essential stuff. It is objective. It answers questions and helps us to solve problems. It helps us to predict. It puts decisions on the right track. Evidence makes sure that decisions are safer. Evidence can turn guesswork into certainty. Evidence tells us what works. It explains why people think and act as they do. It alerts us to likely consequences and implications. It shows us where and when to intervene. We have robust methods for using evidence. Evidence is information; information is abundant. It is the most reliable basis for making policy. Evidence is the most reliable basis for improving practice. There has never been a better time for getting hold of evidence.

Now, what about truth in any of these statements?

Evidence is dangerous stuff. Used unscrupulously it can do harm. It is easily misinterpreted and misrepresented. It is often inconclusive. Evidence is often insufficient or unsuitable for our needs. We will act on it even when it is inadequate or contradictory or biased. We ignore or explain away evidence that doesn't suit our prejudices. We may not spot where evidence has flaws. It can conceal rather than reveal, confuse rather than clarify. It can exaggerate or understate what is actually known. It can confuse us. Evidence can be manipulated politically. We can be persuaded to accept false correlations. A forceful advocate can distort what the evidence actually says.

The answer is that each statement in each cluster is sometimes true, in particular circumstances.

CONTINGENCY

This is not much help to busy people with serious responsibilities who need to act decisively, who need to deliver benefits, value for money, impacts, targets and outcomes. Or busy people who want to get re-elected, reappointed, promoted: ambitious people want to make a mark, leave a legacy, bolster their reputation, earn recognition. Similarly, there are plenty of people who toil away loyally and conscientiously behind the scenes, giving no thought to headlines or sound bites, and who are dedicated to using evidence to improve the quality and fairness of their efforts and those of their service or enterprise or community. What help are the statements to them? Should we conclude that it is impossible to deal wisely with evidence, because evidence is so contingent on circumstances?

On the contrary, this characteristic of evidence is helpful: it encourages us to learn how to be decisive and at the same time to keep an open mind. It obliges us always to be vigilant and sceptical. It reminds us of the hazards of putting too much trust in any one piece of evidence in isolation from other evidence. It counsels us to strive for accuracy in our explanations of what the evidence is telling us.

Recognising three more characteristics of evidence, in addition to contingency, also helps policymakers and practitioners to deal wisely with evidence to improve an existing policy or practice, or to introduce a new initiative to fill a gap. The three are: attributing causality, time lag and good practice.

ATTRIBUTING CAUSALITY

An idea, a practical activity, a policy, a piece of legislation, an international treaty, are among the many elements that may contribute to a specific social or economic change that is valuable (or harmful). Which element(s) caused that change? We want to know so that we can repeat (or prevent) similar effects in future. A claim of causality has to be backed by evidence if it is to convince others, typically evidence of outputs or outcomes.

For example, the Freedom of Information (FOI) Act 2000 gave the UK public a new statutory right to access information held by public authorities and obliged those authorities to publish certain information about their activities. The resulting disclosure and publication of evidence in this way may encourage or sometimes embarrass public authorities into changing particular policies and practices, especially if the media report the story.

One high-profile instance occurred in 2009 when *The Daily Telegraph* obtained and published evidence of MPs' expenses claims, which led to repayments totalling over £1 million, some resignations, a few criminal convictions, and new rules and supervision of claims. The initial evidence came to the newspaper via a leak, which the paper followed up with FOI enquiries.

The line of causality seems obvious at first glance: the leak led to an FOI request which led to disclosures of previously unpublished information. This embarrassed the authorities, the media coverage generated much interest among the general public, emboldening journalists and other observers to call for punishment of the wrong doers and reform of the expenses system, both of which happened.

However, is this explanation the only one? The most accurate one? We must ask whether the equivalent result could have been achieved without *The Daily Telegraph's* actions? Or without FOI legislation? Probably. The authorities might have responded to relentless and increasing media pressure by disclosing the details of the expenses claims anyway. It was not *The Daily Telegraph's* actions that caused the punishments, resignations and reforms to happen. It was the actions of MPs, the parliamentary authorities and the law, against a backdrop of popular interest, one part of which was *The Daily Telegraph's* campaign. Searching for variant explanations that expose more of the texture and interplay of factors help us to avoid jumping to over-simple judgements or seizing on superficial prescriptions.

In most instances of social and economic policy and practice, attributing causality is not straightforward. This is the case irrespective of the specific field or sector, or whether the evidence is quantitative, from, say, large statistical data sets, or qualitative, from, say, service users' views. Attributing causality is so tempting, and leads some people to claim stronger links than are actually in place, especially where the causality seems, to them, so plausible. Consider the example of a policy for culling badgers to reduce bovine TB, a problem that involves many aspects of social and economic policy, scientific inquiry and analysis, and highly-contested interpretations of evidence.

Defra statistics record a rising recent trend in the incidence of TB in cattle in Great Britain, with over 34,000 animals slaughtered in 2011, although the total number of cattle is declining.¹ Cattle get TB from infected cattle and from infected badgers. Randomised controlled trial evidence suggests that (reactive) culling of badgers in and around the farm where cattle had TB increases the levels of TB in that area, while (proactive) culling reduces the incidence of TB within the cull area but increases it in the immediately surrounding area. This is interpreted to indicate that culling changes badgers' territorial behaviour. Other evidence indicates that any benefits of culling are not sustained after the culling ceases.² Alternatives to culling badgers include vaccinating badgers and increasing biosecurity controls on farms, which are also of contested efficacy.

In early 2012 the Government concluded that badger culling trials were justified and set the preparations in motion, but later that year postponed the commencement in the light of further advice that the trials would be ineffective. The President of the Zoological Society and 30 scientists criticised the design of the cull in an open letter, arguing that:

“...the government predicts only limited benefits, insufficient to offset the costs for either farmers or taxpayers. Unfortunately, the imminent pilot culls are too small and too short term to measure the impacts of licensed culling on cattle TB before a wider roll-out of the approach. The necessarily stringent licensing conditions mean that many TB-affected areas of England will remain ineligible for such culling.”³

Defra’s Chief Scientific Adviser and Chief Veterinary Officer responded:

“Government policy is based on sound analysis of 15 years of intensive research. Critics are not able to cite new scientific evidence or suggest an alternative workable solution for dealing quickly with this rising epidemic. Culling is just one of a range of measures the Government is taking to arrest the increase in new bovine TB cases, including intensifying testing to remove infected cattle, tighter cattle movement controls, guidance to farmers on stopping badgers on contacting cattle and further research into vaccination.”⁴

Many other interested parties participate vigorously in this continuing contest for evidence and its interpretations, including many established campaigning and lobbying groups as well as ad hoc coalitions of farmers, countryside and wildlife organisations, the media and individual citizens.⁵ This and the FOI example remind us that ‘what works’ is always a reflection of the context in which policies are constructed, their content and methods, and the needs, motives and perceptions of the people claiming the attribution.

A few words on impact, proxies and failure may help at this point, before looking at time lag and good practice.

RESEARCH AS EVIDENCE

Impact, as an indicator of benefit (or harm) associated with an action or policy or research project, has become an increasingly prominent measure of individual and organisational performance, much favoured nowadays by governments, funding agencies and in the public services. In some respects this an understandable (and somewhat late) response to the recognition that value for money matters. But what exactly are impacts? How are impacts best measured? What sort of evidence is relevant to impacts? These seemingly simple questions are not so simple to answer, and give rise to much friction and disagreement. Accountability for delivering value from public money means identifying what ‘value’ different interests are seeking. Target regimes (for hospital waiting lists, for re-offending rates, for example) are meant to focus planners and policymakers and front-line staff on specific priorities, and may use incentives or penalties to encourage compliance. We know that they also prompt gaming and displacement.

As the ‘value’ question can be so complex, assessments of impacts often turn to proxies. One proxy for academic research quality and effectiveness is article publication and citations in peer-reviewed journals. This proxy is widely used by research councils and higher education funding agencies to finance universities. Money is another proxy for impact. For example, the British Library, like many other publicly funded bodies and services, felt it needed harder evidence to support its claim that its existence and activities were causing valuable economic benefits for the UK’s economy. In 2003 it published evidence from a study it had commissioned, which calculated that for every £1 of public funding the Library received each

year, it generated £4.40 for the economy; and that if public funding of the Library were to end, the UK would lose £280 million per annum.⁶ School students' test results are a further proxy example, taken to indicate the effectiveness of teaching in schools.

In other words, the mere existence of a role, team, department or service, is no longer sufficient justification for automatically continuing former levels of investment in those people and organisations. Nor are their outputs alone regarded as sufficient guarantee of worth or success in delivering value, whether measured in, say, numbers of FOI requests or exam grades or articles published or administrative actions or policy processes. Impact evidence is also demanded.

Because of negative associations to failure, much more evidence is available claiming that something works and much less about what underperforms or fails.⁷ Failure and success are parts of a spectrum of performance. Some proxies for impact are used to argue that policies or activities are failing or underperforming.⁸ For example the percentage of women and members of black and minority ethnic groups recruited to some work roles has been increasing, but not to anything like the extent that many regard as essential.

Here too, the use of evidence to attribute failure or underperformance of a policy or practice depends on the vantage points of the interested parties making that assessment. Re-evaluation of past policies invokes hindsight, and brings different knowledge, expectations and capabilities into the judgement. This can sometimes be a deliberate 'versioning' of the past to promote a current objective.⁹ Analysing policy failures or underperformance can provide opportunities for learning lessons about how to do things differently in future, although some of the cautions mentioned below may apply, on the likelihood that good practice will be adopted.¹⁰

Impact has become a favoured element among paymasters who want to see evidence that will reassure them that they can answer positively the question: 'What difference did your investment in that particular policy/public service/practical action/research make? If they can point to evidence that impact is happening, they feel their investment is legitimised.¹¹ Similarly, critics want evidence that will fuel the accusations of failure they seek to make.

This elevation of the role of evidence of impact is embodied, for example, in the Alliance for Useful Evidence's own title and in its aim to ensure that "...high-quality evidence has a stronger impact on the design and delivery of our own public services." The impact of social science research can be categorised as follows: instrumental (for example, influencing the development of policy, practice or service provision, shaping legislation, altering behaviour); conceptual (for example, contributing to the understanding of these and related issues, reframing debates); capacity building (for example, through technical/personal skill development).¹²

A coalition of seven third sector organisations, calling themselves 'Inspiring Impact', got together in 2012 to develop a programme that will "ensure every pound spent makes the biggest possible difference to beneficiaries."¹³ It aims over the next ten years to:

- Create a code of good practice and practical guidance.
- Create an impact measurement diagnostic.
- Make data, tools, systems and approaches more accessible.
- Explore common indicators and tools for specific fields or interventions to help share and compare results, methods and lessons, and identify the most effective solutions.
- Encourage funders to embed impact in funding decisions and build in evaluation costs and help funders measure their own impact.¹⁴

TIME LAG

There can be a significant time lag between making a decision about or acting on a particular policy or practice, idea or research finding, and seeing the evidence that the result (desired or unwanted) has followed as a direct consequence. Evidence about the effects of some decisions and actions can only be seen (or appreciated) years or generations after the decisions or actions were taken.¹⁵ There are very many examples of this, including taxation and benefits policies and their effects on unemployment; using custodial sentences to reduce crime; changing the classification and penalties for using certain illegal drugs to influence drug-taking behaviours; altering land use planning to encourage specific types of urban or rural development; or introducing new curriculum and examination regimes for schools to improve educational attainment.

The attribution of causality in such cases is complicated for two reasons. First, many factors commonly contribute to a (desired or unwanted) change, not all of which can be identified at the time or later, whether or not these can be reliably measured. Second, the time lag introduces new circumstances and conditions, possibly unforeseen originally, which may play a significant part in the subsequent course of events and thus affect the accuracy of attributing cause and effect. The new factors may, for example, conceal previous effects, or reverse them, or multiply them. An example is the claim that the creation of the European Union, a huge social, economic and political initiative, has prevented wars in Western Europe since 1945.

GOOD PRACTICE

One kind of evidence that many people consider relevant is 'good practice' (sometimes 'best practice'), which often originates elsewhere. Good practice appeals to those who want to avoid re-inventing the wheel locally, to save time, money and effort. They hope that the good practice will equip them to spot potential pitfalls to avoid in advance and identify efficient ways to resolve problems that might well occur. They may also feel more confident about embarking on adopting the good practice because they believe the inventor's own experience with it has already supplied some kind of test of roadworthiness.

Inventors of good practices are very numerous. Individuals and organisations, whether service users or providers, policymakers or practitioners, researchers, teachers or commentators, proudly claim improvements have flowed from their own particular innovations and approaches. If they are keen advocates, articulate enthusiasts, happy to explain what they did and how successful it has been, their proselytising may generate impact points for them too. They may be officially rewarded for spreading the word.

For example, the Northern Rock Foundation was established in 1998 when the Northern Rock Building Society demutualised and became a plc. It has developed its own impact evaluation method and applied it to regional projects it has funded within its 'Safety and Justice Programme'. It obtained the necessary academic expertise through a Knowledge Transfer Partnership (funded by ESRC and the Technology Strategy Board) with the University of Bristol. It is currently promulgating its approach to others as it believes this will have wide applicability to other trusts and foundations and to those who apply for their funding. It has published a number of resources online and actively presents the information at meetings.¹⁶

Yet the record of good practice uptake continues to be rather limited. Why is the scale of uptake so much less than it would be if the majority of eligible beneficiaries adopted it? The answer is a mixture of three things: the 'not invented here' attitude, opportunity costs,¹⁷ and ownership.

- a. 'Not invented here'. Pride, competitiveness, defensiveness and fear of failure can all fuel the rejection of someone else's ideas. Many of us are reluctant to admit that the originator's skills or knowledge are stronger, smarter, better than our own, or to admit that we have been too slow off the mark, or complacent, blinkered, unimaginative, missing the opportunity to be the inventor ourselves.
- b. Opportunity costs will arise because a potentially useful good practice needs preliminary investment of time and money to learn more details from the inventor, to assess what is involved in adapting the ideas or practicalities to local needs and circumstances,¹⁸ perhaps even to run a trial. This may involve efforts to build local interest and commitment. Even then, approval of the change may be the prerogative of others, and may not be forthcoming if they have other pet schemes to roll out, or if they regard the political costs of adopting good practice from the particular source as too high.
- c. Ownership. The identity and rank of the person leading the attempt to adopt a change also affect what happens. Attempts to impose change from above are likely to encounter powerful resistance, especially among those who have to make the change work day by day, unless they have been brought in to the decision early on and feel like part-owners of the change.¹⁹ Similarly, grass-roots attempts to build upon imported or home-grown good practice will struggle if other people in the team or organisation have not been asked to help shape the process of adoption. Authority is not enough; blind obedience is rarely sustainable. Nor does genuine enthusiasm or sweat of the brow suffice. Consent and commitment underpin the sense of ownership and these do not usually occur spontaneously. Participation, engagement, consultation, explanation are the basics that have to be built, they are not optional extras, and they take time and thought to get right.

EVIDENCE AND FACTS

Facts are usually regarded the best kind of evidence. Morgan defines facts as 'shared pieces of knowledge' that are 'short, specific (non-abstract), and reliable. They are non-conjectural: they are not hypotheses, theories, fictions, etc. Nor are they matters of mere belief or opinion. They are established according to criteria of evidence existing in a community at a given time. Facts and other kinds of evidence can be used in new contexts.'²⁰

Some people and organisations consider, for example, that evidence generated by university researchers and published in academic journals has higher trustworthiness and reliability than evidence from, say, newspapers or consultants' reports, or that randomised controlled trials are the 'gold standard' method for generating high-quality research evidence.²¹ Whereas some other people take evidence presented in mass media or blogs seriously.

Even where some facts are available, gaps or inconsistencies in a collection of evidence can persist, for example if reliable knowledge has not yet been created or is not yet understood or sufficiently developed, or not yet found, or if access to it is restricted, as some of the examples above illustrate. Other forms of evidence may be all that is available, even if these too contain gaps or inconsistencies. A decision or action may nevertheless be regarded as valuable by some of the interested parties, perhaps 'better than nothing', even where the shortcomings are known; whereas others may regard that same policy as unsafe or even damaging. Methodological limitations too can hamper evaluations of what works.

According to Gilligan, a sociologist, “the problem with evidence-based policy is not the existence of evidence, or expertise, but the way in which evidence is employed as a substitute for debate.”²² Gilligan has examined this in relation to migration policy, arguing that the current state of policy discussion around migration attempts to treat it as an issue of management²³ rather than a matter involving principles or goals, which could and should be publicly debated. He proposes four principles to inform that debate, but he also acknowledges that it is difficult to prescribe actions or arrangements that would ensure the relevant institutions take note and change their ways.

To sum up, evidence is bound to disappoint those who want conclusive proof from it. Evidence alone does not ensure wisdom or deliver something called ‘objectivity’ or ‘the truth’. Evidence alone cannot quickly silence doubts (about climate change and the role of renewable energy sources, for example). Nor does evidence settle once and for all the value of a specific activity or policy (such as support for SMEs to drive economic growth). Evidence is always contingent on context, sources, perceptions and timing. Good evidence may be ignored, bad evidence may be used misleadingly. Knowing all this helps us to use evidence wisely.

ENDNOTES

1. House of Commons Library 'Bovine TB Statistics: Great Britain.' Standard Note SN06081, 22 October 2012, p8.
2. House of Commons Library 'Badger Culling.' Standard Note: SN 5873, 23 October 2012.
3. 'The Guardian.' 14 October 2012.
4. 'The Times.' 17 October 2012.
5. See also the discussion in Grant, W. (2009) Intractable policy failure: the case of bovine TB and badgers. 'British Journal of Politics and International Relations.' 11, 2009, pp557-73.
6. See: <http://www.bl.uk/aboutus/stratpolprog/increasingvalue/publicvalue/confpres/pungelwesmarks.pdf>
7. Dalziel, M. et al., (2012) Impact of government investments in research and innovation: review of academic investigations. 'Optimum Online.' 42/2, June 2012.
8. McConnell, A. (2010) 'Understanding policy success: rethinking public policy.' Basingstoke: Palgrave Macmillan; and (2010) Policy success, policy failure and grey areas in-between. 'Journal of Public Policy.' 30/3, December 2010, pp345-62.
9. Perche, D. (2011) Dialogue between past and present: policy evaluation and history. 'Australian Journal of Politics and History.' 57/3, pp.403-18.
10. See for example, Skogstad, G. (2007) 'Policy failure, policy learning and policy development in a context of internationalization.' Presentation to workshop on policy failure, Canadian Political Science Association Annual Meeting Saskatoon, Saskatchewan May 30, 2007; Dods (2010) Why policies fail: the ten main risks to successful policies. 'Westminster Explained.' 2010. See: http://www.westminster-explained.com/fileadmin/westminsterexplained2010/images/Policy_Portal_Files/Policyguidance/Why_policies_fail_-_The_ten_main_risks_to_successful_policies.pdf.
11. See for example two reports by the British Academy: (2008) 'Punching our weight: the humanities and social sciences in public policy making.' London: British Academy. September 2008; and (2010) 'Past, present and future. The public value of the humanities and social sciences.' London: British Academy. June 2010. See also a series of publications by Research Councils UK, including 'Impacts: success in shaping public policy and services.' Swindon: Research Councils UK. Available at: <http://www.rcuk.ac.uk/Publications/reports/impacts/Pages/PolicyandServices.aspx>. See also the ESRC's 'How to maximise impact.' <http://www.esrc.ac.uk/funding-and-guidance/tools-and-resources/impact-toolkit/what-how-and-why/how-to.aspx>; and Tinkler, J. (2011) 'Impact for Social Sciences and the Handbook for Social Scientists.' London: LSE Public Policy Group, 21 June 2011. See: http://www2.lse.ac.uk/LSEHealthAndSocialCare/impacts/impact_2011/21_june_presentations/tinkler.pdf
12. See: http://www.esrc.ac.uk/_images/Taking%20Stock_tcm8-4545.pdf
13. See: <http://inspiringimpact.org/>
14. See: <http://inspiringimpact.org/our-plan/>
15. "...while policy looks forwards, evidence looks back." Stilgoe, J. (2007) Three words we should like to hear: 'I don't know.' 'Times Higher Education.' 19 January 2007.
16. Lilley, S.J. and Hester, M. (2012) 'Adding value to grant-making: Northern Rock Foundation's 'Research/Project' Approach.' See: <http://www.nr-foundation.org.uk/wp-content/uploads/2012/11/Exec-Summary.pdf>. Presentation to NCVO/VSSN Researching the Voluntary Sector Conference, 10 September 2012.
17. See also a list of 'wicked' problems in: Wilson, A. and West, S. (2012) 'RCUK proposed research programme on connected communities.' AHRC, 20 July 2012.
18. On international policy transfers see: Fawcett, P. and Marsh, S. (2012) Policy transfer and policy success: the case of the gateway review process (2001-10). 'Government and Opposition.' 47/2, pp162-85.
19. See also Cairney, P. (2012) 'How can policy theory inform policy making (and vice versa)?' Paper to Political Studies Association conference, 2012; see also a special issue of the journal, 'Evidence and Policy' with articles on different patterns of evidence production and use 'Evidence and Policy.' 7/1, January 2011.
20. Howlett, P. and Morgan, M.S. (Eds.) (2010) 'The nature of evidence: how well do 'facts' travel?' Final report 2004-2009. London: LSE. P5.
21. Puttick, R. (2011) 'Ten steps to transform the use of evidence.' London: NESTA. P5. See also Haynes, L. et al., (2012) 'Test, learn, adapt. Developing public policy with randomised controlled trials.' London: Cabinet Office.
22. Gilligan, C. (2012) Immigration controls and the erosion of popular sovereignty. 'GRITIM Working Paper Series.' No.10, Winter 2012. One non-academic observer called immigration control policy "...an example where the policy discourse is 'evidence free'".
23. The International Organisations for Migration says the aim of migration governance is to improve and foster the development of policies that promote humane and orderly migration for the benefit of all. 'World Migration Report 2010.'

Alliance for Useful Evidence

1 Plough Place
London EC4A 1DE

www.nesta.org.uk/areas_of_work/alliance_for_useful_evidence

Alliance.4UsefulEvidence@nesta.org.uk

November 2013