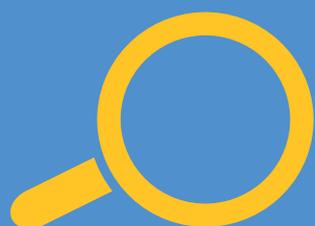
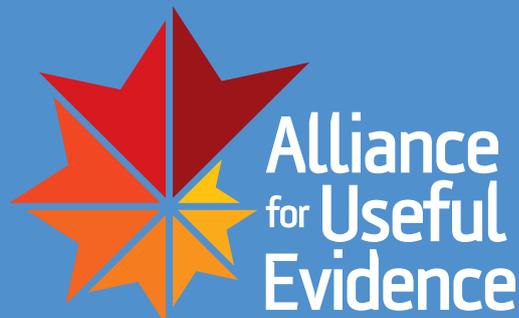


Nesta...



USING RESEARCH EVIDENCE

A Practice
Guide

ACKNOWLEDGEMENTS

This guide was written by **Jonathan Breckon**, edited by **Isobel Roberts** and produced by Nesta's Innovation Skills team.

Thank you to the following for their comments, suggestions and help with this guide: **Helen Anderson** (Policy Profession, Civil Service Learning); **Jonathan Bamber** (Nesta); **Albert Bravo-Biosca** (Nesta); **Professor Paul Cairney** (University of Stirling); **Michael O'Donnell** (BOND); **Triin Edovald** (Nesta); **Caroline Kenny** (Parliamentary Office for Science and Technology); **Geoff Mulgan** (Nesta); **Tony Munton** (Centre for Evidence Based Management and RTK Ltd); **Nicky Miller** (College of Policing); **Helen Mthiyane** (Alliance for Useful Evidence); **Jonathan Sharples** (Education Endowment Foundation); **Louise Shaxson** (Overseas Development Institute); **Howard White** (Campbell Collaboration).

We would also like to thank all the participants and partners in our Evidence Masterclasses during 2014-15, including Civil Service Learning, ACEVO (Association of Chief Executives of Voluntary Organisations), Scottish Council for Voluntary Organisations, SOLACE and the Welsh Government Policy Profession.

The views do not, however, necessarily reflect the suggestions made by these individuals or their affiliated organisations. Any errors in the report remain the author's own.

About the Alliance for Useful Evidence

The Alliance champions the use of evidence in social policy and practice. We are an open-access network of 2,500 individuals from across government, universities, charities, business and local authorities in the UK and internationally. Membership is free.

To sign up visit: alliance4usefulevidence.org/join

Nesta's Practice Guides

This guide is part of a series of Practice Guides developed by Nesta's Innovation Skills team. The guides have been designed to help you learn about innovation methods and approaches and put them into practice in your own work.

For further information contact: skills@nesta.org.uk

About Nesta

Nesta is an innovation charity with a mission to help people and organisations bring great ideas to life. We are dedicated to supporting ideas that can help improve all our lives, with activities ranging from early-stage investment to in-depth research and practical programmes.

Nesta...



USING RESEARCH EVIDENCE

A Practice Guide

CONTENTS

| | |
|-------------------------------------------------------------------------------------|----|
| INTRODUCTION | 4 |
| SECTION A: What is evidence-informed decision-making, and why focus on research? | 6 |
| SECTION B: When can evidence help you? | 13 |
| SECTION C: 'Horses for courses' – What evidence should you choose? | 18 |
| PART 1: Different types of research methods, designs and approaches | 20 |
| PART 2: How do you judge the quality of research? | 29 |
| SECTION D: Where should you look for evidence? | 38 |
| SECTION E: How should you communicate your findings? | 46 |
| SUMMARY | 49 |
| APPENDIX 1: Rationales for evidence and types of evidence required for policymakers | 50 |
| APPENDIX 2: Experimental designs for evaluating complex interventions | 51 |
| APPENDIX 3: Guidance on evidence quality, designs and methods | 52 |
| ENDNOTES | 54 |

INTRODUCTION

“There is nothing a government hates more than to be well-informed; for it makes the process of arriving at decisions much more complicated and difficult.”

John Maynard Keynes

Research evidence can help you understand what works, where, why and for whom. It can also tell you what doesn't work, and you can avoid repeating the failures of others by learning from evaluations of unsuccessful programmes.

Evidence also challenges what we might think is common sense. For instance, it may sound like a good idea to increase the amount of police on the streets to reduce crime or to reduce classroom sizes – but the evidence doesn't necessarily support this. More uniformed police patrolling the streets might make the public feel safer, but it can actually take police away from solving crimes.¹ Despite this, the majority of political leaflets and manifestos in the 2015 UK General Election still claimed that increasing police numbers on the street would reduce crime. Politicians ignored the evidence.²

When budgets are tight, we can't afford to waste money on policies and programmes that don't work. Getting laptops and high-tech gizmos into every school may be a good headline-grabbing commitment, but it does little in itself to benefit young people's learning. In fact research by the OECD found that frequent use of computers in schools is often connected with lower academic results.³

Throwing money at most problems will of course often do some sort of good. But could that money have been spent more effectively on something cheaper? It seems obvious that reducing classroom sizes would improve teaching and learning quality by boosting the one-to-one attention learners receive. But overall the evidence shows little benefit. To achieve impact, typical classes would need to be reduced to 15 pupils – a costly undertaking for those managing school budgets. Instead there are cheaper and higher impact alternatives, such as giving feedback teachers or peer tutoring.⁴

Whether it's in a police station, a school classroom or the boardroom of a charity, evidence can help you make better decisions. It is helpful not only in frontline service-delivery, but also in creating smarter organisations – charities, local authorities, government departments – and in developing national policies or charity campaigns.

We have created this guide to point you on the right path to finding what evidence might help you. It should help to build your confidence in 'compiling, assimilating, distilling, interpreting and presenting a strong evidence base'⁵ using existing research, and to also think about how you might go on to evaluate your own work.

Who might read this guide?

This guide is aimed at decision-makers in government, charities, voluntary organisations, professional membership bodies and local authorities working in UK social policy and practice. It is not aimed at trained evaluators and researchers, but instead intends to foster intelligent demand for research evidence from wider audiences.

How to use this guide

The guide is divided into the following five sections:

SECTION A: What is evidence-informed decision-making, and why focus on research?

SECTION A discusses what we mean by evidence-informed decision-making, and why research is an essential element of it.

SECTION B: When can evidence help you?

SECTION B explores different scenarios in which using evidence can help you, as well as the types of evidence you might need at different stages of development.

SECTION C: 'Horses for courses' – What evidence should you choose?

SECTION C looks at different types of evidence and examines how to choose the most appropriate for your case. It also discusses how to judge the quality of evidence.

SECTION D: Where should you look for evidence?

SECTION D offers advice and resources to help you find the right evidence to support your case.

SECTION E: How should you communicate your findings?

SECTION E focuses on how to get your message across once you have the evidence you need.

SECTION A

What is evidence-informed decision-making, and why focus on research?

This section discusses what we mean by evidence-informed decision-making, and why research is an essential element of it.

To begin, let's be clear about what we don't mean. We are not talking about slavishly following rigid research conclusions. Professional judgement and other sources of information – such as feedback from your stakeholders – will always be important. This guide is not about replacing professional judgement, or pretending that there are easy answers, but about increasing the quality of evidence use in social policy.

A good start in defining what we mean is borrowed from medicine. Two decades ago, David Sackett and his colleagues proposed the following definition that has stood the test of time:

“Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.”⁶

This attempt to define Evidence-Based Medicine (EBM) was not the first, but it has been influential and is just as relevant to social policy as it is to medicine. It stresses how research can complement professional judgement or other sources of information.⁷

However, any model of good decision-making should be wary of putting professional judgement on a pedestal as experts can sometimes get it horribly wrong. Later in this section you will read about how we can be ‘predictably irrational’ and – consciously or unconsciously – make errors in important judgements. In medical practice, [one study](#) found that ‘cognitive factors’ contributed to 74 per cent of mistaken clinical diagnoses – leading to injuries and death.⁸ We will explore how to mitigate these errors of judgement in later chapters.

Other decision-making models have also stressed the importance of blending knowledge of evidence with judgement. The UK Civil Service⁹ recommends that UK policymakers should have skills on ‘politics’ and ‘delivery’ to complement their ability to marshal and apply evidence. This is also relevant to others outside policymaking, such as staff in charities or local authorities. The politics might be slightly different – involving the small ‘p’ politics of your sector or institution instead of manifestos and political parties – but it is still something to master. Delivery is important for any organisation, but the focus of this guide is on bolstering evidence.

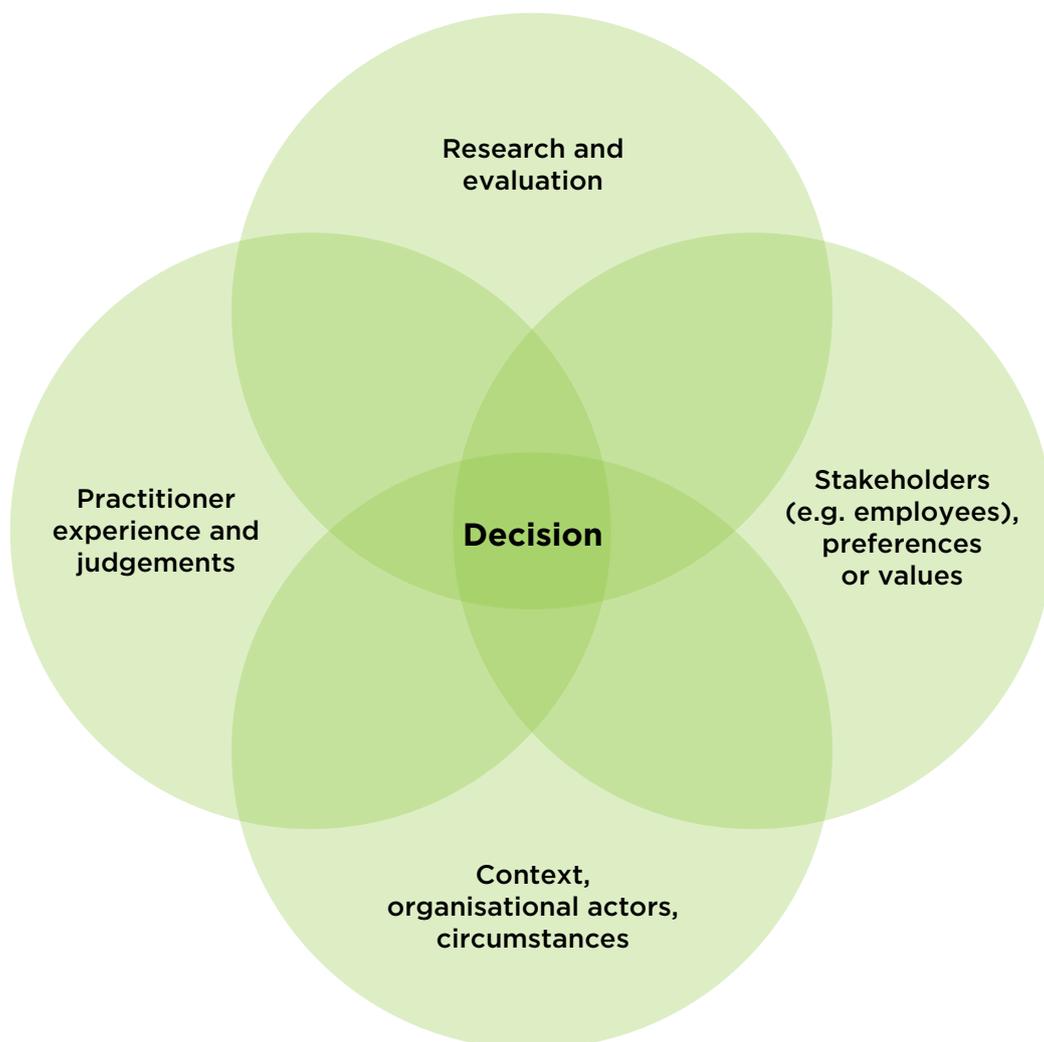
What is 'evidence' and why do we focus on research?

The Oxford English Dictionary definition of 'evidence' is: *"The available body of facts or information indicating whether a belief or proposition is true or valid."*¹⁰

We use this dictionary version because many other definitions tend to be rather unhelpful by being overly inclusive – including almost all information – or by being too abstract and vague.

However, we depart from these definitions somewhat by zeroing in on just one type of evidence. Figure A.1 shows the different elements that should be part of evidence-based decision-making, but our focus is on the top circle of the Venn diagram: research and evaluation.

Figure A.1: **The four elements of evidence-based management**



Source: Professor Rob Briner, Centre for Evidence-based Management

The reason for this focus is nicely captured by an earlier guide on evidence use produced during the creation of the ESRC UK Centre for Evidence Based Policy in the early 2000s:

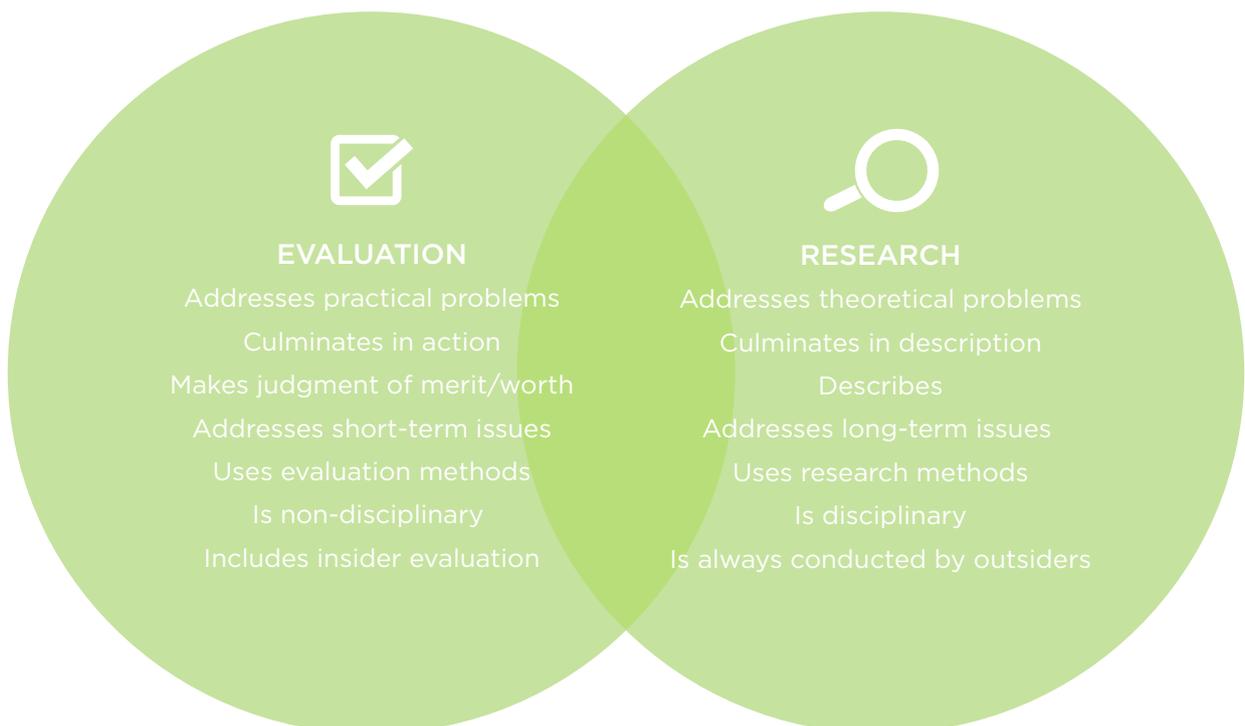
“When we refer to ‘research evidence’, this includes evidence from published research articles and papers, or unpublished sources such as internally conducted evaluations. Research is only one sort of evidence, but has the advantages of greater **rigour, relevance and independence** when compared to some other types of evidence.”¹¹

So why does research evidence have these advantages? As the authors of the Alliance for Useful Evidence report *What Counts as Good Evidence?* state:

“The conduct and publication of research involves the explicit documentation of methods, peer review and external scrutiny, resulting in rigour and openness. These features contribute to its systematic nature and help provide a means to judge the trustworthiness of findings. They also offer the potential to assess the validity of one claim compared to another.”¹²

Other types of evidence – such as in-house evaluations – can be useful and sometimes have these traits. But ‘internal evaluations’ can be hard for others to locate; they are often hidden on organisations’ websites and may not have been peer-reviewed. Too many internal evaluations are also of poor quality (see page 29), and are more likely to be subject to the many biases that can afflict research and evaluation (see page 11).

Figure A.2: Common traits of evaluation versus research¹³



This guide focuses on research, but there are many overlaps with the field of evaluation and we discuss some approaches to evaluating impact in Section C. We also give most attention to research that deals with impact – whether something has had positive or negative results – as questions on impact are vital to our audience. They are concerned about showing their ‘impact’ in charities, ‘results’ in international development or ‘what works’ for government. The language may change, but the ideal stays the same: to see if you have really made a difference.

We also give prominence to research and evaluation that is ready-made, with no need to run a brand-new study. Many decision-makers are pushed for time and resources and simply can’t afford to commission new research. The need for evidence is often now, not in a year’s time, so decision-makers require material that can be taken ‘off the shelf’. Fortunately, this is possible and we cover this in Section D.

Experts vs dart-throwing monkeys: the trouble with professional opinion

Another reason we privilege research is because professional judgement can fail to get it right. A study by the American psychologist and political writer Professor Philip E. Tetlock¹⁴ found the predictions of most experts were wrong. He gathered 80,000 expert predictions and compared them to what actually happened. The results were devastating. Academics, government officials, journalists and other pundits performed worse than ‘dart-throwing monkeys’ in forecasting the future. Indeed, those specialists who had more detailed subject knowledge seemed to perform even worse than average.

Cognitive bias

But experts are not just bad at predicting the future. They can also make more immediate mistakes. A key insight from psychology is how often people make errors of judgement – not just lay people, but highly-trained surgeons, judges or forensic scientists working in life-or-death scenarios. We carry a ‘heavy burden of prejudices, preconceptions and even partiality’.¹⁵ As humans we are ‘predictably irrational’¹⁶ and may experience up to 150 cognitive biases that distort our thinking. Cognitive bias refers to our inability to be entirely objective, which may manifest itself via several possible routes – such as perceptual distortion, inaccurate judgments and illogical and/or irrational interpretations. It means that we should be highly cautious about the accuracy of expert decisions.¹⁷

Even with the best intentions, professionals can get it very wrong. Take for instance one type of cognitive error, confirmation bias. This is the tendency to see the evidence that fits with what we believe, and to ignore or discount what doesn’t. Even highly experienced experts can fall into this cognitive trap (see the case study on page 10).

A study of social workers assessing the risk of child abuse ‘at the front door’ found their accuracy to be ‘only slightly better than guessing’. Confirmation bias was one of the factors to blame for this difficulty in decisions on child abuse, according to the Behavioural Insights Team who conducted the study for the UK’s Department for Education.¹⁸



FORENSIC SCIENTISTS FALL INTO THE TRAP OF 'CONFIRMATION BIAS'

The 2004 train bombings during Madrid's rush-hour killed 191 people and wounded 1,800. The Spanish authorities suspected al-Qaeda. In the hunt to find the bombers, and under the shadow of 9/11, three FBI fingerprint experts confidently concluded that a print taken from a bag containing detonation devices belonged to Brandon Mayfield, an American lawyer in Oregon.

Mayfield spent 17 days in FBI custody. This was despite constant pleas from Spanish authorities that it couldn't possibly be him. Even after finding out that Mayfield's print was not an identical match to the print left on the bag of detonators, FBI fingerprint examiners still rationalised away the differences. Many pieces of evidence were cherry-picked to support their thinking. Some were not-so-forensic, such as the fact that Mayfield had converted to Islam after marrying his Egyptian wife.

Following this embarrassing episode, the US Department for Justice ordered a full review of the case and ultimately implicated 'confirmation bias' as contributing to Mayfield's misidentification, adding that a "*loss of objectivity*" led examiners to see "*similarities... that were not in fact present*".

The FBI had repeatedly twisted facts to fit its theory that he was guilty. In fact the FBI's belief that it had its man, despite all contrary evidence, was so strong that it provided misleading sworn statements to a judge.¹⁹

Being aware of how we can jump to conclusions is important for making us wary of experts. But confirmation bias also highlights how anybody – not just experts – can be highly selective in their use of research evidence. We look for the evidence that fits our beliefs, intentionally or not. This is sometimes referred to as 'policy-based evidence': cherry picking and retro-fitting the evidence to the conclusions we have already fixed in our minds. If we are generous this can be an unconscious mistake, and at worst a deliberate ploy to back up our prejudice by finding the evidence that fits and burying the stuff that doesn't. The Nobel prize-winning psychologist Daniel Kahneman, along with colleagues Dan Lovallo and Olivier Sibony, set out 12 ways to tackle confirmation bias in a useful article in the Harvard Business Review.²⁰

A similar but different phenomenon is optimism bias. This is a very real danger for anybody setting up a big new project in the public sector or charities – particularly one involving new buildings, civil engineering, IT and outsourcing. One influential study of 258 transport infrastructure projects worth US\$90 billion found with overwhelming statistical significance that the cost estimates in planning projects are 'highly and systematically misleading'.²¹ We often focus too much on the positive, which can lead to major cost overruns.

COGNITIVE BIASES RELEVANT TO LEADERS IN SOCIAL POLICY AND PRACTICE



Optimism bias/Planning fallacy

More likely to falsely expect positive outcomes when planning for the future.



Confirmation bias

Tendency to interpret and search for information consistent with one's prior beliefs.



Hindsight bias

Tendency to see past events as being more predictable than they were before the event occurred.



Loss aversion

Tendency to prefer avoiding losses than to acquire gains.



Framing effect

Drawing different conclusions from exactly the same information presented in different ways (e.g. would you prefer a ready meal that's '85 per cent fat-free' or '15 per cent fat'?).



The 'availability heuristic'

When people relate the size, frequency or probability of a problem to how easy it is to remember or imagine.



The 'representativeness heuristic'

When people overestimate the probability of vivid events.



The 'need for coherence'

The urge to establish patterns and causal relationships when they may not exist.²²



Meta-cognitive bias

The belief that we are immune from biases!

(with thanks to Professor Rob Briner, Bath University and Professor Paul Cairney,²³ University of Stirling)

This is not to say that professional judgement is always wrong. Other researchers such as Gary Klein have sung the praises of intuitive expert judgement in his work on ‘naturalistic decision-making’. Professional views and gut-instincts can be highly valuable. But we must be aware of their downsides. As Daniel Kahneman asserted in a joint article with Gary Klein in *American Psychologist*, their influential research has flagged that “*professional intuition is sometimes marvellous, and sometimes flawed*”.²⁴



THE TRAGEDY OF PROFESSIONAL IGNORANCE OVER COT DEATH RESEARCH

Cot death is a horror that can haunt new parents. Fortunately instances of cot death, or to use the medical terminology Sudden Infant Death Syndrome (SIDS), have gone down. But one of the biggest of tragedies of cot death is that if we had looked at the research instead of listening to the experts, many babies’ lives might have been saved in Europe, the US and Australasia.

Following the advice of health professionals such as the best-selling Dr Spock, a whole generation of parents laid their babies face-down in the cot believing that they were doing the right thing. But according to Dr Ruth Gilbert of University College London in an article in the *British Medical Journal* in 2008, by 1970 there was significant evidence from clinical research that putting babies to sleep on their front increased the risk of cot death compared with putting babies on their backs.²⁵ The ‘Back to Sleep’ public health campaign had a dramatic effect on sudden infant death, but was not launched until November 1991, and the safer sleeping position was not consistently recommended until 1995.

Dr Gilbert believes that the advice to put infants to sleep on their front for nearly half a century was “*contrary to evidence available from 1970 that this was likely to be harmful*”. A systematic review (see page 34) of preventable risk factors for SIDS from 1970 would have led to earlier recognition of the harm of sleeping on the front and might have helped prevent some of the 10,000 infant deaths in the UK and 50,000 in Europe, the US and Australasia.

KEY MESSAGES FROM SECTION A

- We are not advocating slavishly following research at the expense of professional judgement – or other sources of information – but **you also shouldn’t put professional expertise on a pedestal**. Research has shown that experts can sometimes get it horribly wrong, and they are not immune from a whole range of social and cognitive biases.
- Creating new evidence can be costly and time-consuming, but there is plenty of **good quality evidence already in existence** that can be taken ‘off the shelf’.
- Research is only one sort of evidence, but has the advantages of **greater rigour, relevance and independence** when compared to other types of evidence.

SECTION B

When can evidence help you?

This section explores different scenarios in which using evidence can help you, as well as the types of evidence you might need at different stages of development.

Evidence can make organisations more effective. From more persuasive campaigning to winning grant funding; and from developing a board's decision-making abilities to making sure programmes deliver results – evidence can bolster your work. It doesn't matter if you are a small voluntary organisation or a large government department. Whatever the scale, there's a type of affordable research that can suit your needs.

Below are some examples of when evidence can help you:

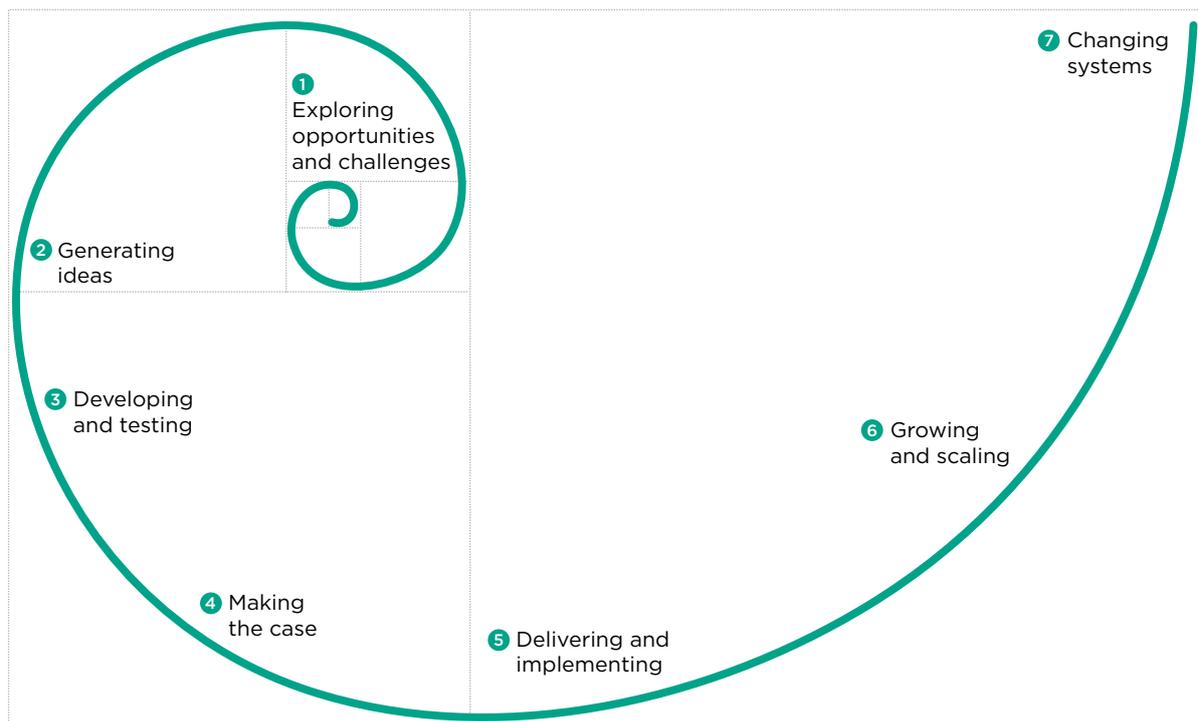
Figure B.1: Reasons for needing evidence



It's a good idea to begin by think about timing. Different evidence will be helpful at different times. You will need to think about the appropriate research to suit different stages of the lifecycle of a new programme, policy or practice.

In the early days of a new initiative, research can identify emerging challenges and the scale of the problem. For instance, a longitudinal study (see Table C:2 on different research approaches) observes people over time – sometimes decades – may highlight the deteriorating public health of a large group of people, such as the depression and morbidity associated with loneliness in old age, or the health and psychological damage of physical inactivity in children.

Figure B.2: The Nesta innovation spiral



The spiral above was developed by Nesta to capture the different stages of the innovation process, and can also be used to plot the progress of a new approach to a social issue.²⁶ Different types of evidence will be needed at the different stages:

1. Exploring opportunities and challenges

Working through current research will help you to understand the problems and opportunities around your particular area or issue.

2. Generating ideas

After you've identified your focus it's time to hunt for evaluations of interventions and policies that have worked – or failed – in the past. Can you borrow successful ideas from others and avoid interventions that have failed?

3. Developing and testing

As new ideas are initiated, it will be time to start thinking about testing and experimenting with different approaches – and about evaluating the impact of these. In section C, we discuss Standards of Evidence that show the journey new programmes should go on in terms of providing evidence of impact.

4. Making the case

Having evidence of the impact of your testing will then help with making the case to funders and supporters. This will put you in a stronger position to persuade your backers and move on to delivery and implementation.

5. Delivery and implementation

Once you have implemented a policy, programme or project you need to think about how you can evidence your impact. This will help take you from a position of rhetoric and saying 'trust me, this project is working' to one based on more trustworthy evidence. It will also help with your accountability – by showing that you are making a difference, that you are value for money, and have opened up your work to evaluation.

6. Growing, scaling and spreading

The holy grail for many innovations is to replicate and grow so that they change things on a bigger scale – and can potentially be successfully copied in other locations (see Nesta's report *Making It Big: Strategies for scaling social innovations*²⁷).

7. Changing systems

Achieving system-wide changes is extremely complex, but can be seen in the mainstreaming of recycling, the hospice movement or the ban on smoking in public places. A good way to reach this scale is to have multiple independent studies to show that something is not just a flash in the pan but an effective approach that works in many places.

Adopting ideas from other regions or sectors

Even if you find evidence of success of a policy or practice, would adoption of those ideas work in your area?

For instance, there seemed to be plenty of positive research to support the Nurse Family Partnership (NFP), a US-created programme that uses specially trained nurses to carry out home visits with teenage first-time mothers from low-income families. It has shown some astonishing success, with over two decades worth of evidence to back up its effectiveness. But much of that research comes from the US, so would it work in other countries?

It certainly doesn't seem to be the case in the UK. An evaluation found that NFP didn't seem to translate and wasn't as effective in the UK – perhaps because through the NHS, British mothers already had more support than US ones.²⁸ This example shows that context matters, and that we should not cut-and-paste programmes from elsewhere, unless they are adapted to local circumstances.

And when designing new programmes aimed at changing how we act – for example, cutting smoking or reducing anti-social behaviour – it's always worth checking out research from those in the fields of social psychology and behavioural research, such as The Behavioural Insights Team. There is a range of techniques to 'nudge' people towards desired actions based on robust research such as randomised controlled trials or ethnographic research (see Section C on research methods and approaches).

CREATING A THEORY OF CHANGE

In the early stages of any intervention, it's important to logically describe what you do and why it matters coherently, clearly and convincingly. This is often referred to as a Theory of Change, and aims to give a:

“...clear, concise and convincing explanation of what you do, what impact you aim to have, and how you believe you will have it. It is a vital foundation of any programme, and a prerequisite for effective evaluation.”²⁹

A Theory of Change is a useful way to be more explicit about what evidence you are using from others – and to be clearer about how you are going to get results. For instance, if you want to run a programme to get kids more physically active, what evidence is out there on effective behavioural change? Is building more playgrounds enough? What incentives really work on children (and their parents) to get them off the couch? You could get some useful pointers from behavioural science or the evaluation of other physical activity programmes.

A Theory of Change helps you be explicit about your goals – and how you'll achieve those goals. It helps to avoid just 'winging it', hoping that your new innovation may stick and that your assumptions are correct. Another benefit of doing a Theory of Change is that it's a first step in designing an effective evaluation as it tries to accurately identify all of your outcomes that then need to be measured.

For more on policy development, rather than growing charities or programmes outside government, see Appendix 1 for rationales for evidence and types of evidence required to match your policy questions.

The most important message here is to think about appropriateness. You want to find the research that fits your needs and the stage of development that you are at. The sort of evidence you'll require in the first few months of a new programme will be very different to what you need when you are more established. We revisit this crucial issue of appropriateness in the next section.

KEY MESSAGES FROM SECTION B

- There is a **wide range of situations in which evidence can help you**. It's not just the obvious ones around capturing the results of programmes, policies and practice. There may be other benefits that you haven't thought about, such as creating more persuasive campaigning, winning grant funding or stopping doing something that's not working.
- You need to think about the **timing and appropriateness of evidence**. Think where you may be on the innovation spiral – and what evidence may suit your needs. For instance, in the early days of developing a new innovation you'll most likely want to avoid doing a large, costly evaluation of your own work and instead learn from others through past evaluations or wider social and scientific research.
- Early stages of innovation are also the time to create a **Theory of Change**. This is a useful way to be clearer about what evidence you are using and how you are going to get results. It will also help you to design an effective evaluation.
- Programmes, policies and practices that are **more established can set up their own experiments or evaluations of impact**. As an innovation grows, you should also consider multiple replications to check that success was not just a flash in the pan, but that it can work in other places and contexts.
- **It doesn't matter what scale of organisation** you are. There is now a lot of freely available research even if you are a small voluntary organisation.

SECTION C

‘Horses for courses’ – What evidence should you choose?

This section looks at different types of evidence and examines how to choose the most appropriate for your case. It also discusses how to judge the quality of evidence.

Not all evidence is equal. Some is stronger – and more relevant to your challenge – than others.

A key message of this practice guide is to think about appropriateness. Too many people get into trouble by not thinking clearly about what sort of research design, method or way of collecting data is really going to meet their needs. We can be ‘methods-led’.³⁰ In other words, we pick our pet approach, such as questionnaires or RCTs, because that’s what we’re most comfortable with rather than being more open to the breadth of methods – and thinking which ones are best suited to answer our challenge.

For instance, experimental research is more suited to evaluating impact and ‘what works’. The UK network of ‘Sure Start’ children’s centres – which provide help and advice on child and family issues – were informed by randomised control trial evaluations and systematic reviews (see page 23 for more on these approaches) of early years initiatives in the US.³¹ But other types of research can help give other insights. For instance, policymakers setting up Sure Start were inspired by longitudinal studies – research tracking people over decades – that showed the long-term trends of children’s poverty.

Other sorts of research may not look so much at impact, but can reveal why and how things are working.³² For instance, Sunderland City Council found that using ethnography³³ – where the researcher observes from the point of view of the subject – helped Sunderland redesign its employment and welfare programme. The council was able to get closer to its customers’ needs by putting itself in the shoes of citizens through ethnography. They were then able to make major savings by getting citizens back into sustained work (see case study on the following page).



LEARNING BY DOING - ETHNOGRAPHY AND WHAT WORKS IN FINDING A JOB

Back in 2007, more than a quarter of the working age population in Sunderland were economically inactive and Sunderland City Council was spending a significant amount on benefit claims. But very little progress was being made, so the council tried a new approach; they carried out ethnographic research to unpick the real story of people's often difficult and bureaucratic journey back to work. Ethnography allowed them think about services from a user perspective: through travelling with and talking to people on that journey, they were able to design a suite of services that supported them through it.

In its initial phase the Make It Work programme supported more than 1,000 people, generating early savings of more than a quarter of a million pounds for the council. As well as using ethnography, their approach was experimental, informed by rigorous evidence and tested by iterative prototyping – checking if things were working, and learning quickly and cheaply from mistakes and successes.

What this example stresses is that we don't always know what works – there are gaps and uncertainties in the evidence. But we can take an experimental approach and test as we go, rather than unrealistically thinking there will be a fully-formed solution ready for implementation.

The programme's success depended on Sunderland Council's willingness to try something new using evidence based on local user needs, to fund it properly and to give it the space and permission to experiment. The total cost of running the programme was £180,000. An economist for the council has estimated overall cost avoidance for the council of £435,000 through participants entering sustained work, amounting to an initial saving of approximately £255,000.

This early return on investment is dwarfed by the long-term savings of reducing worklessness. The current Minister of State for Welfare Reform at the Department for Work and Pensions has estimated that it is economically rational to spend £62,000 on supporting the average unemployed person back into work. The average cost of participation in Make it Work is only £5,000.

SOURCE: Radical Efficiency; different, better, lower cost public services, Nesta 2010

Whatever the method, whether it is trials or ethnography, the type of research needs to fit the needs of the challenge – in other words: 'horses for courses'.³⁴

PART 1: Different types of research methods, designs and approaches

So how do you go about understanding which research method is appropriate for your case? If you are not a researcher, it can be daunting to come face-to-face with the jargon and endless lists of approaches.³⁵ See for instance these examples:



Adapted from Luff *et al.* (2015)

But such long lists shouldn't put you off as you don't need to know every type of approach. For the non-specialist 'consumer' of research, it's more important – and easier – to understand the assumptions that underpin these ways of doing research.³⁶

These assumptions are captured by a Department for International Development (DFID) *How to Note*³⁷ on assessing evidence – and are just as relevant to social policy as international development:

- Some research designs are better suited for **demonstrating** the presence of a causal relationship, such as experimental and quasi-experimental designs.
- Others are more appropriate for **explaining** such causal relationships – see Beyond experiments: new approaches to evaluation on page 26.
- While some designs are more useful for **describing** political, social and environmental contexts – such as observational studies.

To understand a bit more about the pros and cons of these approaches, see Table C.1.

Table C.1: **Different designs, methods and approaches to research evidence – a brief overview**

| Types of research and evaluation | What is it? | Pros | Cons |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Personal anecdote | An account by a person with direct experience of, or affected by, a particular issue. | Powerful and immediate; may give vivid insights into events concealed from much of the population. ³⁸ | Difficult to verify; may lead to inflation of prevalence; emotive first-person stories may inhibit critical appraisal; individual anecdotes may not be representative. |
| Service use feedback | Narrative or survey accounts of user views or reported needs. | Valuable insights from those at the receiving end; compels professionals to stay focused on the service users' priorities. | Correlation between satisfaction and service effectiveness is low; expressed needs may not translate into actual service use. |
| Single case study | Detailed and intensive examination of a single case such as community, family, organisation, event or individual. | Easy and practical; can be used by practitioners and non-specialists; good for interventions that have already happened; can possibly identify adverse consequences from intervention; helps to describe innovations; generates hypotheses for other studies. | Bad at inferring causations; small size means hard to generalise to national/ population level. |
| Case control studies | Compares a group who have, say, a health condition with a group of people that do not have it, and looks back in time to see how the characteristics of the two groups differ. | Require fewer resources to carry out than intervention studies; useful when RCTs not practical, e.g. studies of cot death. | Rare in social policy (see closely related 'case-based' evaluation below for approach more common in social policy), more frequent in epidemiology and health; provide less evidence for causal inference than an RCT; high risk of bias e.g. recall bias, selection bias, interviewer bias. |
| | | | |

| | | | |
|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Participatory | An approach where the judgements and experiences of stakeholders and beneficiaries are highlighted. May be described as normative designs; participatory or democratic evaluation; empowerment evaluation; learning by doing; policy dialogue; collaborative action research. | Beneficiaries are best able to identify the most relevant theories of change and meaningful outcomes; more potential to be ethical and democratic and understand what beneficiaries really need; more appropriate to 'values-led' interventions; more opportunities for programme adoption due to closer ties to beneficiaries; ability to adapt and customise intervention. | Argument that it is fundamentally un-objective, open to bias and not really research but more about ideology and community activism. ³⁹ |
| Theory-based | An approach to evaluation that looks at what happens compared with pre-existing theories or causal pathways identified during an evaluation. Can be associated with realist evaluation; qualitative comparative analysis (QCA); contribution analysis; process tracing. | Strong on explanation of causes; can be used in messier areas of social policy where there may be many causes and context is important. <i>[Note: this category is used very broadly here, for more detail on this area see Stern (2015)⁴⁰ and White and Phillips (2012)⁴¹].</i> | Lack of agreed methods; opportunities for bias; weak on estimating quantities or extent of impact; relatively little evaluation of the techniques used e.g. compared to the large body of literature criticising RCTs. |
| Cross-sectional study | A representative sample of people surveyed at one point in time. Although surveys such as questionnaires and structured interviews are commonly used in cross-sectional design, they are not the only way e.g. content analysis or analysis of official statistics can be used. | Quantitative data can be examined to detect patterns of association; relatively cheap and ethical; survey can be repeated at intervals, illustrating changing trends over time (see Longitudinal below). | Establishes association at best, not causality; rarely possible to attribute any measured change to the intervention, or to understand what would have happened in the absence of the intervention – e.g. change could have been to broader issues such as economic conditions, weather, media campaigns – and not the intervention. Other disadvantages: risk of recall bias, social desirability bias, researcher's (Neyman) bias; group sizes may be unequal; confounders may be unequally distributed. |
| Cohort/ Longitudinal studies | The same sample of people surveyed over several points in time, sometimes from childhood to old age. | Best source of evidence on association between childhood experience and adult outcomes can give powerful support for certain early interventions. | Data often emerges too late for effective policy-making; study members may drop out over time; very expensive approach when maintained over decades. |
| | | | |

| | | | |
|------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Quasi-experimental design</p> | <p>Different interventions are offered but with no random allocation to groups, i.e. through the use of natural populations or case matching.</p> | <p>Can provide reasonably strong evidence of the relationship between the intervention and the measured outcomes; powerful method of exploring the impact of an intervention when randomisation is impossible; can be applied to large communities as well as groups; no need for randomisation from the start (ex-ante), which avoids the PR and ethical issues of randomisation.</p> | <p>Inability to ensure equivalence of groups and to prevent change over time can result in less reliable findings; matching techniques tend to require a lot of data in both intervention and comparison groups which can be time-consuming and expensive; a good understanding is required of the factors that need to be matched – without this, it remains possible that there are systematic differences between the two groups that are not being controlled for; these designs require complex analytical work and specialist knowledge.</p> |
| <p>Randomised control trial (RCT)</p> | <p>One group receives an intervention while another receives none or one of another type, with the chance of trial (RCT) being allocated to either group being identical.</p> | <p>Offers the most robust, reliable findings, which give confidence that any measured difference between groups are the result of the intervention; random allocation should overcome any systematic difference between groups; greater confidence in the effect size, and the relationship between the intervention and outcome; internationally recognised approach.</p> | <p>Poor on taking context into account e.g. cultural, institutional, historical and economic settings; difficult to conduct at a national population level; risk that when used at small pilot level not relevant to national/ population level (although this is a risk for all designs); can be hard to manipulate variables to experiment in social policy e.g. class, race or where you live; mistakes in randomisation can invalidate results; can be hard to persuade decision-makers (e.g. politicians) of benefits of this design; potential political, ethical and PR issues over randomisation (e.g. some groups randomly getting potential beneficial intervention, and not others); can take more management time and longer to set up than quasi-experiments.</p> |
| <p>Systematic reviews/meta-analysis</p> | <p>Aggregation of results from eligible studies, with the eligibility criteria defined in advance and methodologies reviewed.</p> | <p>Best source of reassurance that an intervention works (or doesn't); meta-analysis pools statistical results; large reviews carry considerable statistical power; is replicable by other researchers; can be applied to any kind of data.</p> | <p>Requires a substantial number of robust primary studies in a given area; methodology less well developed for synthesising qualitative data and 'grey' literature [<i>For wider approaches to synthesis, see other approaches such as narrative synthesis, and realist-based synthesis</i>].</p> |

Adapted from: Quality in policy impact evaluation: understanding the effects of policy from other influences. (supplementary Magenta Book guidance) HM Treasury/DEFRA/DECC (HM Government: 2012); The Evidence Guide; Using Research and Evaluation in Social Care and Allied Professions (2006) Barnado's/What Works for Children?/Centre for Evidence-Based Social Services; and Petticrew, M. & H. Roberts (2003), 'Evidence, hierarchies and typologies: horses for courses'. 'Journal of Epidemiology and Community Health.' 57: 527-529 (2003); Stern, E. (2015) 'Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.' Table 2, p18.

Experimental research – why all the fuss?

Experimental research such as randomised controlled trials (RCTs) have received a lot of attention⁴² – some might say too much. So why have they attracted so much interest?

Frequently those working in social policy want to know if an intervention has had an impact. Has there been a causal link between your new programme and the ‘outcome’, ‘impact’, ‘effect’ and ‘result’ at the end? While they are often hard to implement in practice, experimental designs have a somewhat better chance of tracking this cause and effect.

Experiments can simply test two non-randomised groups before and after an intervention. But there may be a risk that the groups are biased or not representative of the groups under investigation – for example, somebody who asks for intervention may be more predisposed to positive outcomes. Perhaps they are more motivated, healthier, confident, and thus more likely to do well regardless of any ‘treatment’.

To get rid of this bias, we can use randomisation: the groups compared with each other are selected entirely randomly, for example by drawing lots. The box below gives a fuller description of this approach, taken from a supplement to HM Treasury’s *Magenta Book*.

EXPERIMENTAL RESEARCH DESIGNS

“An experimental design, conducted properly, will establish whether an intervention caused an outcome. Such evaluation designs use random allocation to assign units of assessment (individuals/groups) to either the intervention or counterfactual group (often called ‘control group’ in experimental design). Given appropriate sample sizes and appropriate allocation to experimental or control groups, this is the strongest form of design for an impact evaluation, as the random allocation minimises the likelihood of any systematic differences – either known or unknown – between the groups. It therefore allows for an attribution of cause and effect.”

Taken from *Magenta Guide: HM Treasury, DECC and DEFRA (2012) ‘Quality in policy impact evaluation; understanding the effects of policy from other influences.’*

The introduction of a control group eliminates a whole host of biases that normally complicate the evaluation process – for example, if you introduce a new ‘back to work’ scheme, how will you know whether those receiving the extra support might not have found a job anyway?



MAKING IT WORK – EXPERIMENTAL RESEARCH ON INCAPACITY BENEFIT

In 2003, the Department for Work and Pensions (DWP) conducted a randomised controlled trial to examine the impact of three new programmes for Incapacity Benefit claimants. These were: support at work, support focused on their individual health needs, or both. The extra support cost £1,400 on average, but the RCT found no benefit over the standard support that was already available. It ultimately saved the taxpayer a lot of money as it provided unambiguous evidence that the costly additional support was not having the intended effect.

More recently the DWP was keen to explore whether the intensity of the signing-on process required of jobseekers on benefits could be reduced without worsening outcomes.

In a trial involving over 60,000 people, the usual fortnightly signing-on process was compared against several others that were less resource intensive (e.g. signing-on by telephone, less frequently). All of the alternatives to the status quo that were tested in trials large enough to show reliable effects were found to increase the time people took to find work. As a result, despite other changes to the benefits system, DWP policy continues to require people to sign on on a fortnightly basis.

Source: Cabinet Office (2012) *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*.

Practical and ethical issues

A common criticism of RCTs is that they are impractical, but there are often ways round this. For example, the so-called ‘stepped wedge’ design (see Appendix 2 on different ways of doing trials) can be used as a solution to some of the ethical and practical problems. In this approach, everybody does, eventually, get the intervention – such as new learning approach in schools or a crime-reduction innovation.

But, with stepped wedge trials, the order in which participants receive the intervention is determined by lottery. It is gradually phased in over time – but who gets it is chosen at random. This can be highly practical in a world of austerity. As there is frequently not enough resources to do a full roll-out of a new programme all in one go, a gradual roll-out can be cost-effective. Appendix 2 has a short overview of this and four other different experimental research methods.

USEFUL GUIDES ON EXPERIMENTAL RESEARCH

- *Randomised controlled trials – gold standard or fool’s gold? The role of experimental methods in voluntary sector impact assessment* – a ten-page basic introduction to RCTs for charities by NCVO/Charities Evaluation Service
- *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials* – a helpful guide from the Cabinet Office
- *Better Public Services through Experimental Government* – a Nesta report on some of the myths around the ethics and costs of RCTs
- *Developing and evaluating complex interventions: new guidance* – excellent guidance on a variety of RCTs and other ways of evaluating complex interventions from the Medical Research Council (with examples from social policy, not just medicine)

When randomisation is not appropriate

Randomly choosing your subjects for research may not be appropriate. It may be completely unrealistic to deliberately expose your experimental ‘subjects’ to something harmful such as cigarette smoking or a crime-ridden housing estate. It may be more appropriate to use non-randomised designs or ‘natural experiments’.

The cot-death case study (see page 12) is an example where nobody in their right mind would run an experiment – with some babies sleeping on their fronts and others on their backs to test if they die or not. But we can still compare groups that have had different experiences. The research on cot-deaths used ‘case controlled studies’⁴³ (see item six in Table C.1).

There are other examples of research that have also bypassed traditional RCTs but still used comparable groups to see if there is some link. For example we have been able to compare different cases on legislation that restrict access to the means of suicide, the impact on air pollution controls in different areas and the impact of banning smoking in public places.⁴⁴

Quasi-experimental designs

There are a wide variety of ‘quasi-experimental designs’ (QEDs), often with cumbersome titles such as Interrupted Time Series Designs, Proxy Pre-test Design, Non-equivalent Dependent Variables Design, Pattern Matching Design and the Regression Point Displacement Design.

The influential proponent of RCTs and QEDs in the 1960s in the US, Professor Donald Campbell, often referred to them as “queasy” experiments because they give experimental purists a queasy feeling. But QEDs are still helpful when RCTs are not feasible. The Alliance for Useful Evidence has worked with the European Commission (EC), JPAL Europe and the LSE to produce *Testing Social Policy Innovations*⁴⁵ that covers many common quasi-experimental designs, available on our website.

Beyond experiments: new approaches to evaluation

A lot of current thinking on impact has moved away from sole dependence on experiments.⁴⁶ RCTs or quasi-experiments may work well when there is a simple intervention that can be tested. However, rarely do we have such simple interventions. NGOs working in civil society, for instance, rarely work alone or have the chance to manipulate a clear experimental ‘treatment’ for an RCT. Evaluators are looking at other ways of doing things, approaches that involve moving beyond talking of simple one-to-one causality – and thinking about your ‘contribution’.

In the complex world of social policy, it’s unlikely that your programme is the necessary or sufficient condition for success. It’s likely to be just one factor among many, part of a ‘causal package’. Programme success depends on what else is going on or has gone on around you. In the past, a simple causal explanation of ‘what works’ may have sufficed but nowadays it is now much more common for evaluation researchers to ask: ‘did we make a difference?’⁴⁷

The case for such theory-based approaches is convincingly set out in an excellent guide produced for the Big Lottery Fund, BOND, Comic Relief and DFID. It shows how explanation, not just causation, is important for any evaluator:

“*You might draw a conclusion (or causal inference) from an evaluation that funding for education programmes for girls led to or ‘caused’ higher family income in a particular community. However, when it becomes evident that similar educational programmes do not always lead to the same result in all places, people start to ask ‘why?’.*

In IE [impact evaluation], as in scientific research, explanation ultimately relies on good theories. Opening up the ‘black box’ that connects ‘causes’ and ‘effects’ requires different kinds of analysis, which is what ‘theories of change’ and ‘programme theory’... are intended to support.

Developments in IE have also made evaluators aware that they need to draw on broader community, social and economic theories in order to interpret complex and often confusing or even contradictory data.”

Source: Stern, E. (2015) *Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector*. Prepared for Bond, Comic Relief, Big Lottery

There are a range of approaches that are becoming more popular with evaluators, such as theory-based⁴⁸ and case-based designs, realist evaluation,⁴⁹ contribution analysis,⁵⁰ process tracing⁵¹ or qualitative comparative analysis.⁵²

Rather than overgeneralising about these methods, it’s worth touching on one example: contribution analysis. This form of research does not attempt to prove that one factor – say a policy – ‘caused’ the desired outcome, but rather to explore the contribution a policy is making to observed results. It recognises that effects are produced by several causes at the same time, not just one. The Scottish Government has a short guide⁵³ on this technique and says that it can be useful when experiments are not practical:

“*By developing a ‘theory of change’ showing the links between the activities, outcomes and contexts of the policy and collecting evidence from various sources to test this theory,*

the aim is to build a credible ‘performance story’. This can demonstrate whether policies and programmes were indeed an important influencing factor in driving change, perhaps along with other factors. ”

These approaches may not all be new.⁵⁴ Indeed, having a good theory has arguably always been at the heart of good science – so ‘theory-based’ approaches don’t sound too novel. But these ways of evaluating impact have grown in popularity and they do help evaluators address multiple causality.

Was it worth it? The value of cost-benefit analysis

If you throw lots of money at any social problem, you are likely to get some benefit. But is it value for money? When budgets are constrained, we need to make hard financial decisions about whether one thing is more valuable compared to another. Policymakers, commissioners of services, grant-making bodies and charitable funders alike are asking for more and better impact measures for every penny they disburse. Techniques such as cost benefit analysis (CBA) can help create such insights on value. It results in a financial value being ascribed to impact, which means that in the current funding climate many charities and public bodies are understandably intrigued.⁵⁵

It’s important to note that there are a range of techniques. A cost benefit analysis (CBA) is only one way to compare costs with financial benefits. Two other approaches are worth mentioning: cost effectiveness analysis (CEA) and cost consequence analysis (CCA). The differences between CEAs and CBAs are quite technical and are one for specialists, but the general difference with CBAs is that they measure results in units rather than monetary figures. Other techniques have also been developed specifically for social and environmental value such as social return on investment (SROI).⁵⁶

Because of this broad range of approaches, if you want to run a CBA on one of your own programmes of work we recommend that you use a tool that is already widely used and accepted – see below for a selection of guides on cost benefit analysis.

GUIDES ON COST BENEFIT ANALYSIS

The HM Treasury’s *Green Book*⁵⁷ – a guide for the appraisal and evaluation of policies, programmes and projects that is relevant for charities and service providers as well as government – features a CBA model frequently used in the public sector.

Organisations wishing to develop this sort of analysis need to collect data on elements such as key performance indicators, costs of service delivery, and (where possible) outcomes or impact on service users. Guidance on what data to collect can be found in another UK government publication entitled *Choosing the Right FABRIC*,⁵⁸ and many organisations operating in the public sector have taken an approach to the collection and analysis of performance data informed by this document.

Another useful guide, particularly for those working on local projects and who have tight analytical budgets, is *Supporting public service transformation: cost benefit analysis guidance for local partnerships* produced by HM Treasury, Public Service Transformation Network and New Economy in Manchester.⁵⁹ It is designed to simplify and lower the cost of performing CBAs for local programmes aimed at improving public services where analytical and research resources may be relatively limited. It’s relevant to charities and business, as well as public sector.

For charities, the Big Lottery Fund’s *A Guide to Cost Benefit Analysis’ Wellbeing Programme: Evaluation and Learning* will also be useful.

KEY MESSAGES FOR PART 1 OF SECTION C

- **Not all evidence is equal.** Some is better quality and will be more appropriate to your challenge.
- Whatever the method, whether it is RCTs or ethnography, the type of research needs to **fit the needs of the challenge** – in other words: ‘horses for courses’.⁶⁰
- Don’t **be put off by long lists of research methods**, designs and approaches. It’s more important to understand the assumptions that underpin these ways of doing research.
- Some research designs are better suited for demonstrating the presence of a **causal relationship**, such as experimental and quasi-experimental designs.
- Other research approaches are more appropriate for explaining such causal relationships. **Theory-based evaluation** and techniques such as contribution analysis are increasingly popular with evaluators. These techniques can be helpful when it’s hard to experiment or impossible to attribute your single policy or programmes to any single clear result.
- It’s not just about whether your intervention worked or not, but whether it was value for money. Use techniques such as **cost benefit analysis** to understand the financial value of your impact.

PART 2: How do you judge the quality of research?

Another way to help you choose which sort of research you need is to ask a different question: what research can you trust? What is good enough evidence to fit your needs? Perhaps it is only the academic top-ranking journals like *Nature*, *BMJ* or the *Quarterly Journal of Economics*. But what about the many in-house evaluations conducted by charities, local authorities and government? Surely they too must have a place on the evidence-table?

Being in a peer-reviewed research journal is one way to help you feel confident of the research, but it's no guarantee. In a famous paper, John Ioannidis from Stanford University caused a stir by arguing that 'most published findings are probably false'. He examined the most cited papers (1,000+ citations) in the best regarded medical journals in the world – largely drawn from *The Lancet*, the *New England Journal of Medicine* and the *Journal of the American Medical Association*.⁶¹ Of those with claims of efficacy whose results had been tested, 41 per cent were either found to be wrong, or the impact was much smaller than the original study suggested. The Ioannidis study is a decade old, but in 2013 he told the International Congress on Peer Review and Biomedical Publication that the problem had not gone away.⁶²

Peer-reviewed research: the decision-maker's comfort zone?

We must also be conscious that peer review – the cornerstone of academic journals – is far from perfect. There can be unconscious biases such as 'herding', where the behaviour of reviewers is influenced by the behaviour of their peers. And the whole setup can be skewed by 'publication bias': positive results have a better chance of being published,⁶³ while negative data gets shelved. A survey by Stanford University found that most 'null studies' in sociology never see the light of day: just 20 per cent had appeared in a journal, and 65 per cent had not even been written up.⁶⁴

This could have serious consequences for decision-makers reviewing evidence. If you never see all the negative studies on a new intervention, you may wrongly conclude that all is well. In the clinical health field, hiding the results of clinical trials of drugs like Tamiflu may literally kill.⁶⁵

Nevertheless, despite these problems, peer review is still, for most, the 'gold standard': a check on bad work.⁶⁶ Using a journal that has been peer-reviewed by other experts is one way of helping you be somewhat more confident that you can trust the research. It still can take years before research gets published and that can be too long if you have to make a quick decision, but peer-review has some quality-controls to help you feel confident.

If the research hasn't gone through a peer-reviewed journal, there may be other channels. For those working in government, it's recommended that you consult learned societies (such as the Royal Society, Royal Statistical Society, British Academy and others) or scientific and expert committees as sensible buttresses of quality.⁶⁷ In the charity sector, bodies like [Evaluation Support Scotland](#) have advised using peer networks and talking to other like-minded organisations.⁶⁸

Nailing down a definition of 'high quality' research

Peer review may give us some modicum of comfort. But what do you do if you are going to include wider evidence that hasn't been checked by other experts? How do you decide if it is up to scratch?

It would clearly be daft to miss important 'grey literature' by research bodies such as the Royal Society for Edinburgh, King's Fund or the Institute for Fiscal Studies just because it wasn't in a peer-reviewed journal. Or to miss the rich seam of government or charity evaluations that never see the 'light' by being published in academic outlets. But it feels a daunting task after reading the National Audit Office's damning report on evaluation.⁶⁹ Looking at central government, it found that less than half of the evaluations they examined had provided robust evidence on policy impact – meaning that the findings could not be relied on.

A good start in trying to vet quality is defining it. One of the problems, however, is that phrases such as ‘quality’, ‘standards’, ‘robustness’ and ‘strength’ are bandied around as if they were interchangeable, and without clearly defining what they are. This makes for a lot of messiness and misunderstanding.

For instance, in some guidance,⁷⁰ research ‘quality’ means using particular types of design and method – such as RCTs. This focus on quality is a deliberate nod to some of the formal clinical and health approaches to assessing evidence quality, such as the GRADE or Scientific Maryland Scale system.⁷¹

But for others, quality can mean good practice in how you report⁷² your research, or your integrity⁷³ in not faking your data – as seen in the academic scandal of Dutch psychologist Diederik Stapel, who fabricated data for years and published it in at least 30 peer-reviewed papers.⁷⁴ Even more confusingly, some researchers have judged quality as being about utility – how relevant your work was to policy and practice.⁷⁵ While utility is a commendable goal, it may sound confusing to non-specialists to equate quality with usefulness.

However, for most of the literature, discussions of research quality have zeroed-in on methodological designs. Particular methods like RCTs have won this privileged position as ‘high quality’. According to a supplement to the HM Treasury’s *Magenta Book*⁷⁶ – an evaluation ‘bible’ – experimental and quasi-experimental designs get closer to seeing if interventions ‘cause’ outcomes (see a short description of these designs in Appendix 2).

HM TREASURY’S MAGENTA AND GREEN BOOK

When thinking about what is strong evidence, it is impossible to ignore two key HM Treasury tomes of guidance – the *Magenta Book*⁷⁷ and *Green Book*.⁷⁸ While both are hugely long – 141 pages and 118 pages respectively – they are readable (all-things-being-relative) and avoid much technical language of economics and evaluation. They are essential reading for anybody making judgements on the strength of evidence.

While both have value for vetting evidence, the *Magenta Book* is particularly helpful in planning your own evaluations or reviewing evidence created by others, whereas the *Green Book* is more future-orientated and perhaps more for officials in central government – offering help on, for instance, how to build a good business case for a new policy.

The *Magenta Book* was also written with wider audiences in mind than Whitehall and so is highly relevant to local authorities, charities, public service providers or external consultants and evaluators that want to work with government.



THE DANGER OF USING ‘WEAK’ EXPERIMENTS TO INFORM POLICY – THE CASE OF SYNTHETIC PHONICS

The current policy for teaching children to read is based on synthetic phonics, by which children learn to match sounds to letters and groups of letters. A team commissioned by the Department for Education and Skills and led by Professor Carole Torgerson, then at York University, was asked to review the evidence on synthetic phonics.

Although there was some promising evidence from the US and a small-scale study in Clackmannanshire, Scotland, showing that the approach worked, the evidence was limited. Their government review found only a dozen small experimental trials, the biggest of which involved 120 children.⁷⁹ They urged caution in making national policy. Torgerson recommended in her report that the government should roll it out gradually, with the first areas to benefit to be chosen at random in a trial.

But this advice was ignored, according to Torgerson, and ‘it just became policy’: “As a result, we still don’t know whether or not phonics works as the main teaching strategy for all children,” she has said. “Some of the recent evaluation work has demonstrated synthetic phonics may not be having the impact that was hoped for. If we’d done a randomised trial we would have known before the policy went national.”⁸⁰

The *Magenta Book* states that there is direct link between quality and research design when thinking of causal questions.⁸¹ High quality impact evaluation is all about the issue of attribution – i.e. can you attribute your intervention as the cause of the outcomes? – and to do this you need a good comparison or control group. Then you can be more confident that your particular good idea improved, say, children’s services, reduced crime, or helped jobseekers back to work.

We should flag here that there are heated and expansive debates about what constitutes causality and attribution.⁸² Many researchers would dispute the privileged position of experimental designs because they are impractical and don’t always tell you why something that works in one place can work in other places.⁸³

This definition of quality may not just be disputed by specialist researchers alone. For anybody outside evaluation, it may seem rather odd to equate quality with experimental designs. It’s perfectly possible to have high quality research in other approaches, such as ethnography, focus groups or online surveys. It’s also just as easy to have a shoddy RCT as in any other type of research.

High quality ‘qual’ research

Many have sung the praises of good quality ‘qualitative’⁸⁴ research for policymakers (see page 19 on Sunderland City Council’s use of ethnography). The well-seasoned evaluation expert Stephen Morris at the independent social research institute NatCen says the most persuasive evidence is ‘rigorous, social science’. He argues that we should not privilege quantitative over qualitative evidence, or vice versa. Qualitative evidence can be just as scientifically credible as quantitative.⁸⁵ But what we should do is:

“...privilege evidence that involves the collection and analysis of data, through the application of defined, systematic and replicable methodologies, producing transparent data, testing predefined evaluation questions, and with results made publically available.”

(Evaluating service transformation, NatCen blog, 19 November 2014).

Using top-tiers, hierarchies and standards of evidence

Another way of approaching this topic of quality is to use formal standards of evidence – particularly when looking at questions of ‘what works’ in social policy and practice.

These can take the form of hierarchies, where stronger research designs such as RCTs and systematic reviews are at the top for ‘causal’ questions – and other methods, usually case studies, are at the bottom.⁸⁶

Figure C.2: Two illustrations of simplified hierarchies of evidence, based on the design of the study

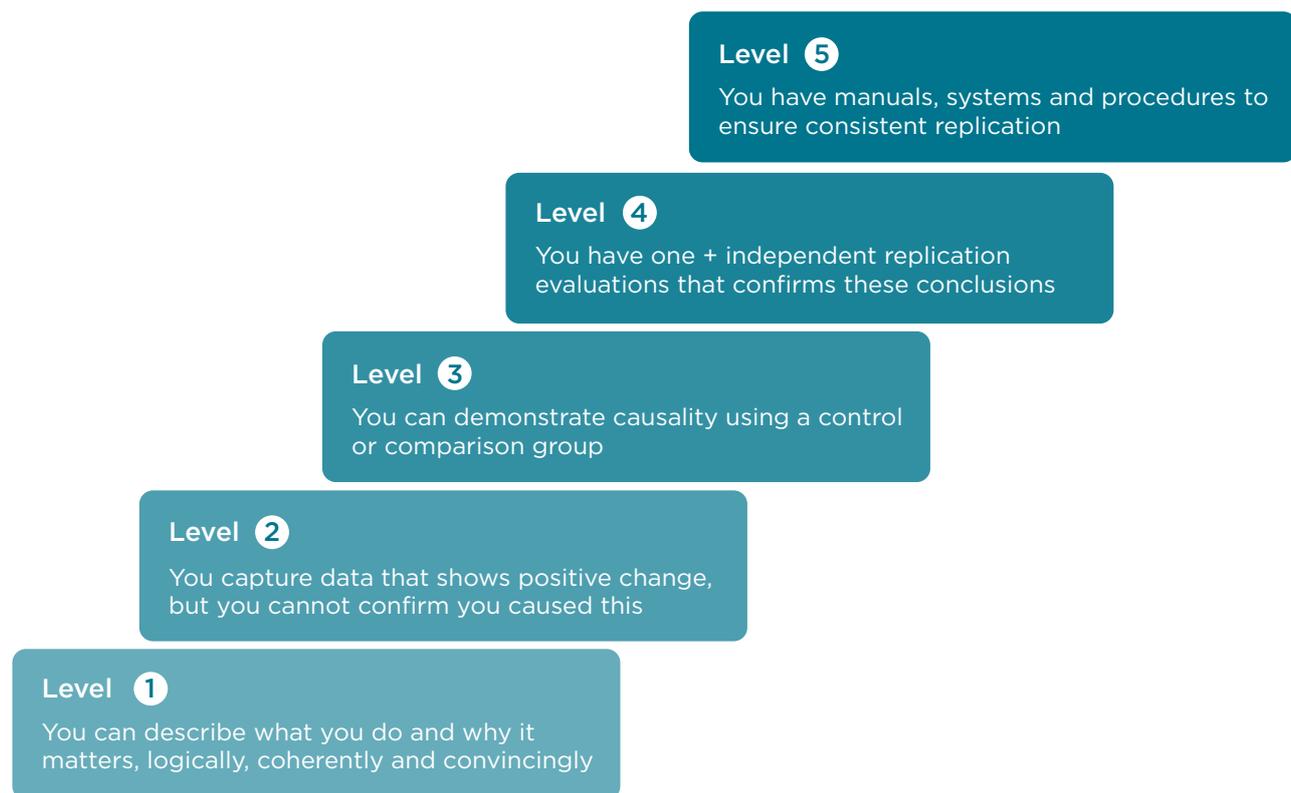


Source: Bagshaw and Bellomo 2008, p2.

Source: Petticrew and Roberts 2003, p.527.

Different versions of evidence hierarchies are used by many evidence review groups and endorsing bodies around the world, and they are particularly prevalent in healthcare. These standards of evidence have also been growing in UK social policy. Bodies that have been using them include Public Health England,⁸⁷ the Nesta/Cabinet Office's Centre for Social Action⁸⁸ and the youth crime reduction Project Oracle in London.⁸⁹

Figure C.3: Nesta's Standards of Evidence



Nesta developed its Standards of Evidence framework in 2012 to assess the quality of evidence on the impact of the social ventures we have invested in. Impact investment is big business and it has been given considerable support by the UK government. But for Nesta's investment fund, the concern was around how we could be sure that the new innovations were really having a positive impact. Could they even be doing harm? Strong evidence was needed to show if real progress was being made with the investments.

The Nesta framework has five levels of evidence.⁹⁰ These start from the basic level (level 1) where you can describe what you do and why it matters, logically, coherently and convincingly. You then move up the scale towards routine data collection (level 2), and then to higher levels where you use comparison groups (level 3), before creating evidence of replication and scaling (levels 4 and 5). The level of evidence you require should be appropriate for where you are in developing a policy or intervention. For Nesta, it's been particularly helpful for the ventures we fund to articulate how their product or service leads to positive social change.⁹¹

Critics of these hierarchies say that they can be overly-rigid straightjackets. Decisions of quality need to change according to context and it shouldn't be one-size-fits-all. Interestingly, the National Institute for Health and Care Excellence (NICE) – one of the UK Government's What Works centres (see page 39) – has now dropped its hierarchy of evidence. Their former Chairman, Michael Rawlins has been outwardly critical of 'slavishly' following evidence hierarchies.⁹² He was in favour of a more nuanced sense of deciding what is appropriate, rather than RCTs and systematic reviews always trumping the rest.⁹³

Whatever the merits of these criticisms,⁹⁴ these hierarchies, frameworks and principles do at least provide a structure to check the evidence claims. The Nesta model has been tried-and-tested. They provide a relatively easy-to-grasp, non-technical structure for the 'evidence journey' that many organisations need to go through; from the basics of a 'theory of change' to multiple replication studies. Evidence standards offer a way of vetting the strength of effectiveness claims, and of avoiding thinking that all evidence is equal.

These standards are also proportionate and take into account early-stage innovations. A new project does not have to show its impact straight away – or commission an expensive evaluation from day one. New ideas need time to grow. But reaching the higher levels of evidence (levels 4 and 5) does beg the question of how you judge the strength of the body of evidence, not just a single study – something we will look at next with systematic reviews.

Avoid cherry picking by using systematic reviews

It's important for decision-makers to think about the quality of the evidence 'base' – not single pieces of evidence, but aggregated collections of research.

Much of the thinking on quality set out above has focused on single studies in 'primary research' – in other words, original research such as experiments, surveys or suchlike. We must also think about 'secondary research', and the summarising and pooling together of existing research – a task that is likely to be desk-based. We are much more likely to find a charity leader or government official undertaking this sort of secondary work.

A common pitfall is to carry out a 'literature review'. These are highly dubious. As Ben Goldacre put it, they encourage the temptation to cherry pick – consciously or unconsciously – your preferred bits of evidence to fit with your preferred conclusions:

“*Instead of just mooching through the research literature, consciously or unconsciously picking out papers here and there that support [our] pre-existing beliefs, [we] take a scientific, systematic approach to the very process of looking for scientific evidence, ensuring that [our] evidence is as complete and representative as possible of all the research that has ever been done.*”

Ben Goldacre (2012) *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients*, (London: Fourth Estate).

We've seen in the earlier section on cognitive biases that we can unconsciously fall into the trap of looking for evidence that fits our beliefs. So we need to be very careful in collecting all the research that is out there, even the research with inconvenient facts.

One way to avoid this cherry picking is to use what are called systematic reviews.⁹⁵ These approaches aim to be exhaustive, and to find as much of the research relevant to your question as possible. They use explicit methods to identify what can reliably be said on the basis of these studies. The reviews can screen studies for quality – on the basis of the research design and methods.

Systematic approaches also have the value of being explicit about how they searched for research – so, in theory at least,⁹⁶ others could replicate the systematic review, unlike the 'mooching around' described by Ben Goldacre above.

CHARACTERISTICS OF A SYSTEMATIC REVIEW

- Has a clearly stated set of objectives with pre-defined eligibility criteria for studies;
- Has an explicit, reproducible methodology;
- Uses a systematic search that attempts to identify all studies that meet the eligibility criteria;
- Includes a formal assessment of the validity of the findings of the included studies;
- Produces a systematic presentation, and synthesis, of the characteristics and findings of the included studies.

Source: HM Treasury (2011) *The Magenta Book; Guidance for Evaluation*

Meta-analysis

Meta-analysis is often an important part of systematic reviews and usually involves bringing together a large amount of research findings – but using statistical analysis. Frequently, it means pooling the average effect sizes estimated in each individual research study.

Meta-analysis is perhaps best known for combining the results of randomised controlled trials, but it is also commonly undertaken on non-randomised data – from primary studies that use case-control to cross-sectional and cohort designs (see Table C1 on page 21).

DISTINGUISHING BETWEEN META-ANALYSIS VS META-EVALUATION VS SYSTEMATIC REVIEWS

There is an important distinction between systematic reviews and meta-analysis, and they should not be conflated. A systematic review uses systematic methods for reviewing literature. It can be considered a ‘meta level’ (about or beyond) process as it is research on research. The term ‘meta-analysis’ could therefore be used to describe all systematic reviews, but in practice the term is used only for statistical meta-analysis, where statistics are used to synthesise the numerical data from different studies.

There is also the term ‘meta-evaluation’ (Scriven 1969) which can mean ‘evaluation of evaluations’, and can include the synthesis of evaluation studies, but also has broader meanings (Scriven 1969, Gough *et al.* 2012b).⁹⁷ For a longer description of ‘meta-evaluation’ and how it is different to other types of reviews, see the *Magenta Book*.

Reviews in a hurry

But systematic reviews and meta-analysis can be time-consuming, which is no good for urgent decision-making. You probably can’t wait around for six months to complete a review in time for a big speech tomorrow, a funding bid deadline or a memo on a new policy for a minister. However, many of these systematic reviews are already ‘in the bag’, so you don’t need to start from scratch. They are freely available in online libraries such as the ones curated by the Cochrane and Campbell Collaborations, discussed in Section D on where to go for evidence.

Nevertheless, you may be frustrated and fail to find what you are looking for on those websites. So your best bet is to commission a ‘pared-down version of the systematic review’⁹⁸ such as a rapid evidence assessment (REA). REAs aim to review evidence in a hurry, without unduly compromising quality (see Section D for more on REAs).⁹⁹

However, since the questions we want answered now are very likely to recur in years to come, at the institutional level it makes sense to commission high-quality reviews that are regularly updated and easily accessible, which is precisely the role of the online libraries covered in Section D.

Note, however, that more formal ways of bringing together research don't always equate to quality. Judging the quality of your evidence review is still hard to do. The quality of your synthesis is only really as good as the quality of the studies it is based on.

The importance of repetition and corroboration

There are other things that need to be taken into account when vetting the quality of a research summary. For instance, how many studies need to be included for you to be comfortable that a strong body of evidence exists? Perhaps only a couple of studies if they are really good? Or maybe dozens, or hundreds? A recent 'Evidence Check' of the Department for Education¹⁰⁰ mentions 12 Ofsted reports on primary schools that purport to show strong performance related to phonics teaching. But there are around 16,000 state-funded primaries in England, so 12 doesn't sound very much. And that's also avoiding the other serious question of whether we can really say that Ofsted inspections count as strong evidence of impact.

The reality is that there is no magic number of studies. Yet we can't ignore the fact that the size of the body of evidence is important: there is strength in numbers, and we must have repetition and corroboration. Even studies that have won many accolades need to be repeated, again and again. Amgen, a Californian drug company, tried to replicate 53 landmark cancer studies. The work, published in *Nature*,¹⁰¹ only replicated six (11 per cent) of these important studies. This is deeply troubling as the studies have influenced drug companies and cancer treatments globally.

'Reviews of reviews' – how to judge high-quality bodies of evidence

We should certainly be very wary of single studies – about new wonder drugs for Alzheimer's splashed on the front page of the *Daily Mail*, for example – that are never replicated.¹⁰² But the size of the evidence is not the only thing to consider. A *How to Note* on judging the strength of evidence produced by the DFID lists four things to consider when checking bodies of evidence:

1. The (technical) quality of the studies constituting the body of evidence (or the degree to which risk of bias has been addressed);
2. The size of the body of evidence;
3. The context in which the evidence is set;
4. The consistency of the findings produced by studies constituting the body of evidence.

Source: DFID (2014) *How to Note: Assessing the Strength of Evidence*

Judging all these criteria is always going to be rather subjective – and relevant to the context of the policy question. It's really hard to give blanket advice for the right body of evidence for any policy. Another interesting approach is to bring together only collections of systematic reviews, and not to sift through just single studies – in other words, a 'review of reviews'. Of course, this may only work if systematic reviews already exist – and in some sectors there are few. A good model is the Education Endowment Foundation – an independent charity and 'What Works Centre' funded by the Department for Education and Sutton Trust, which is discussed in Section D.

For a list of other in-depth guides to research quality, see Appendix 3.

KEY MESSAGES FROM SECTION C PART 2

- To find evidence that you can trust **look for peer-reviewed research**. But note that peer review is far from perfect. There can be unconscious biases such as ‘herding’, or publication bias towards positive results, or even academic fraud.
- When looking at questions of impact and ‘what works’, use the **frameworks and formal standards of evidence** such as those used by Nesta, Project Oracle and others.
- One study is never enough in social policy. **Avoid making decisions based on single studies, and look for multiple replications**. There is strength in numbers, and we must have repetition and corroboration. Even studies that have won many accolades need to be repeated, again and again.
- Don’t do a ‘literature review’ of research, as they encourage cherry picking – consciously or unconsciously – of evidence. **Use systematic reviews**, which aim to be exhaustive and can screen studies for quality – usually on the basis of the research design and methods.
- If you haven’t got time to do a systematic review, or can’t find a completed one that answers your question, commission a **rapid evidence assessment**.

SECTION D

Where should you look for evidence?

This section offers advice and resources to help you find the right evidence to support your case.

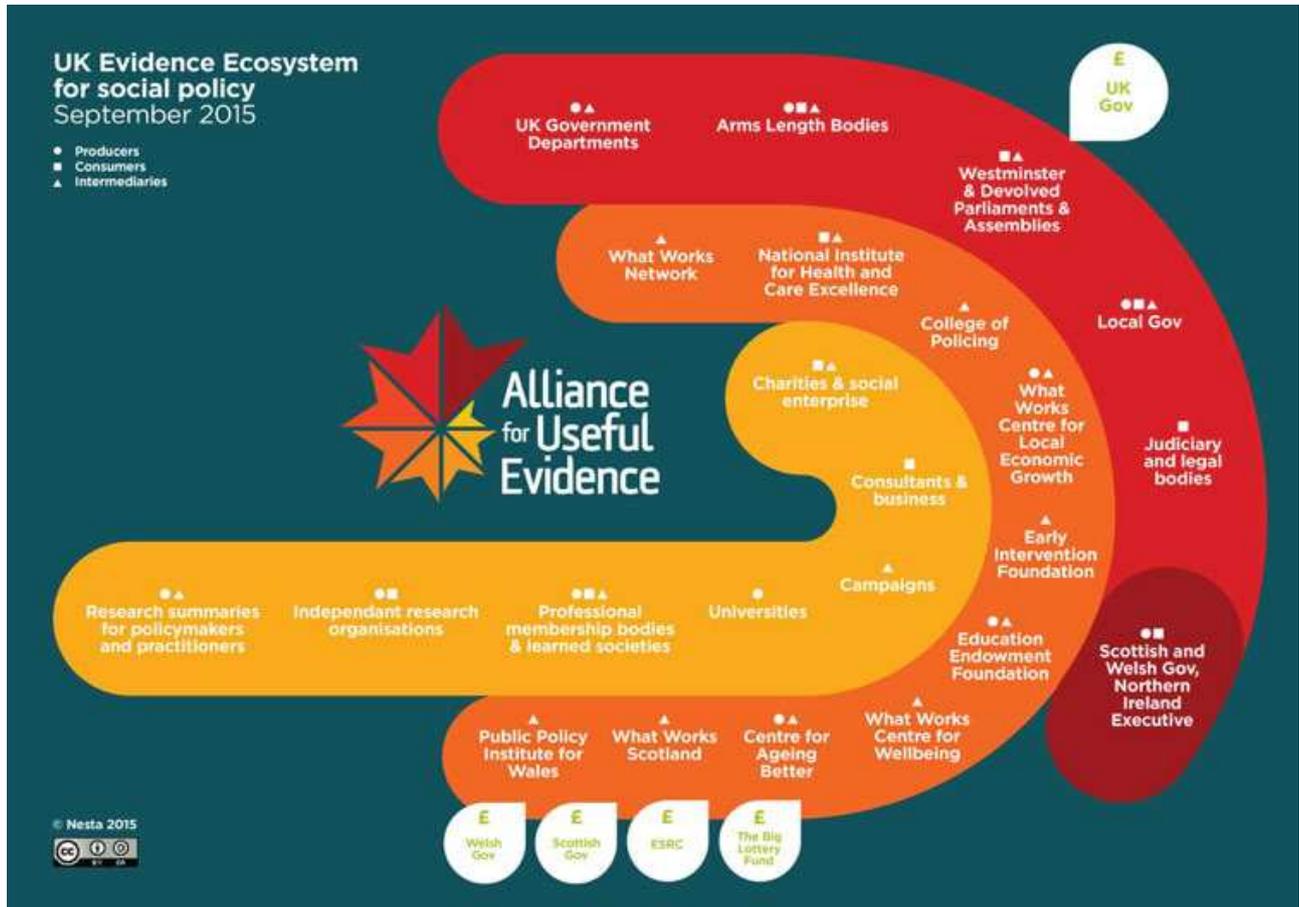
The screenshot shows the Guardian website interface. At the top, there are navigation links for 'become a member', 'sign in', 'subscribe', and 'search'. The Guardian logo is prominently displayed. Below the logo is a horizontal menu with categories like 'UK', 'world', 'politics', 'sport', 'football', 'opinion', 'culture', 'business', 'lifestyle', 'fashion', 'environment', 'tech', and 'travel'. A 'browse all sections' button is also visible. The main article is titled 'Bushfires' and 'Greg Hunt uses Wikipedia research to dismiss climate change-bushfires link'. The sub-headline reads: 'Environment minister warns against 'politicising the science', saying bushfires are simply part of 'the Australian experience''. The author is Helen Davidson, with a Twitter handle @heldavidson. The article is dated Thursday 24 October 2013 04:45 BST. Social media sharing icons for Facebook, Twitter, Email, Print, LinkedIn, and Google+ are present. The article has 6,500 shares and 322 comments. A 'Save for later' button is also visible. A large blue hashtag #GregHuntResearch is overlaid on the right side of the image. Below the article title, there is a photograph of Greg Hunt speaking at a podium with microphones. A small caption below the photo reads: 'Environment minister Greg Hunt told BBC there was no connection between climate change and the severity of the NSW bushfires. Photograph: Alan Porritt/AAP Image'.

It was refreshingly honest of former Australian environment Minister Greg Hunt to admit that he got his evidence from surfing the net. ***“I looked up what Wikipedia says,”*** was his defence on the BBC World Service when quizzed about his government’s views on bush fires. Mr Hunt’s comments went viral. He was ridiculed on Twitter, where his statement spawned the hashtag #GregHuntResearch. To rub home the point, some mischievous Wikipedians updated his personal Wikipedia page with a note that he ***“was quoted as saying he uses Wikipedia for important policy research”***.

You would hope that other custodians of policy would look for more robust research than Wikipedia, but the Australian politician is not alone. A recent survey¹⁰³ by Carnegie UK Trust found that most policymakers accessed evidence through the internet, newspapers and broadcast news.

We are missing a trick here. At the very least, Greg Hunt could have surfed the ubiquitous [Google Scholar](#) search engine. There is a wide range of trusted online research resources, many of them free and easy to access. This chapter will signpost just a few that we think will be useful to any policymaker, charity leader or frontline professional.

Figure D.1: The UK Evidence Ecosystem for social policy



The UK What Works centres in social (and local economic) policy

The launch of six ‘what works’ centres in 2013 has transformed the landscape of evidence for social policy and practice in the UK.¹⁰⁴ Their aim is to improve the way government and other organisations create, share and use quality evidence and they are overseen by the What Works Network run out of the Cabinet Office, as well as the Scottish and Welsh governments. There are now nine centres – seen in the orange central ‘sausage’ in the figure above – covering areas like crime, education and wellbeing. They are funded by the Economic and Social Research Council, the Big Lottery Fund and a whole range of government departments – including the Scottish and Welsh governments.

No longer can we use the excuse that research is too difficult to access because it’s hidden behind academic paywalls, or in some indecipherable jargon. Whether you are in a charity, central government or local authority, they provide accessible and actionable syntheses of research and evidence. A particular priority is to make sure their summaries of research are communicated in a way that resonates with decision-makers – not only policymakers, but also

commissioners of services or frontline practitioners like teachers and police officers. It's still early days for some of these What Works Centres, so not all of their resources are finished yet. But, as we publish, the following are useful and open resources that summarise research:

- **The Sutton Trust/Education Endowment Foundation's (EEF) Teaching and Learning Toolkit** and **EEF Early Years Toolkit** – evidence of impact on pupil attainment for interventions in the classroom or around the school, such as mentoring or sports participation.

Audience: teachers, parents, heads, early years professionals, governors but also any charity, local authority or policymaker.

- **What Works Centre for Local Economic Growth** – systematic reviews on local economic growth policy issues such as access to finance, apprenticeships, broadband, employment training, enterprise zones and transport.

Audience: anyone involved in making policy decisions that are targeting economic growth – especially local authorities, LEPs, government, and businesses.

- **What Works Centre for Crime Reduction Toolkit at the College of Policing** – over 300 systematic reviews, covering 60 different crime reduction interventions, have been identified and will be added to the online tool over time, such as CCTV, street lighting and prison visits to deter young offenders.

Audience: police and crime reduction professionals, policymakers, commissioners of public services, charities and voluntary groups with interest in crime reduction programmes and interventions.

| Intervention | Impact on crime | How it works | Where it works | How to do it | What it costs |
|---------------------------------------------|-----------------|--------------|----------------|----------------|---------------|
| | Effect | Mechanism | Moderator | Implementation | Economic cost |
| Alcohol ignition interlock | ✓ | ⚙️ | 📍 | ? | £ |
| Alcohol tax and price policies | ✓ | ⚙️ | 📍 | ? | £ |
| NEW Alternative education programmes | X✓ | ⚙️ | 📍 | ? | £ |
| CCTV | ✓ | ⚙️ | 📍 | ? | £ |
| Cognitive Behavioural Therapy (CBT) | ✓ | ⚙️ | 📍 | ? | £ |

- **Early Intervention Foundation Early Intervention Programmes Library and Guidebook** – covers the evidence for over 50 early intervention programmes. It looks to encourage positive development and tackle a range of problems, including abuse and neglect, substance abuse and mental health problems. The guidebook is designed to provide advice for professionals developing early intervention programmes and systems.

Audience: commissioners of services, policymakers and practitioners across social policy.

- **Public Policy Institute for Wales' evidence reviews** – the PPIW provides short turnaround analysis and advice for Welsh government ministers. It also undertakes work on some of the key strategic challenges facing the Welsh government, such as tackling poverty. Its website features links to dozens of freely available completed research summaries including rapid evidence reviews, expert advice and analysis by PPIW, ranging from **strategic transport planning** to the impact of indebtedness in Wales.

Audience: Welsh government ministers, others with interest in Welsh social policymaking.

Keep an eye out for forthcoming online summaries of research by the other What Works Centres:

- What Works in Tackling Poverty (led by the Public Policy Institute for Wales – who also run the What Works Centre for Wales).
- **What Works Scotland Evidence Bank.**
- What Works Centre for Wellbeing.
- Centre for Ageing Better.

NICE work in health and social care

A lodestar in the UK What Works centres is the National Institute for Health and Care Excellence (NICE). The guidance offered by NICE is all based on the best available research evidence – as well as inputs from patients through 'citizen jury' type bodies, and clinicians.¹⁰⁵ NICE also offers two other useful resources:

- **NHS Evidence:** a helpful database that gives digestible research and guidance ready-made for the doctor's surgery or hospital ward. It's an extraordinary resource and ranges from open and free **databases on medicines** to more advanced (but unfortunately only for NHS staff) **links** to top academic journals and bibliographic databases.
- **UK DUETS:** short for the Database of Uncertainties about the Effects of Treatments, it flags up where serious question marks remain about whether treatments work or not. It's based on what patients and clinicians are saying, as well as the best available primary research and **systematic reviews**. The database is candid about our gaps in understanding and recognises that 'ignoring treatment uncertainties puts patients at risk'.

Other research databases for social policy and practice

Most databases are not as practitioner-friendly as NHS Evidence, and are aimed more at specialist researchers. Some of these other databases can, however, be surprisingly easy to access:

- The **Evidence for Policy and Practice Information (EPPI)** Co-ordinating Centre at the University College London has a range of databases that are free and online, such as the **Database of Education Research**, which has 1,200 records of systematic reviews. You can search all its systematic reviews by keyword.
- There are also large scientific databases of journals, such as **PubMed** for studies in public health or **JSTOR** for studies in economics, sociology and public policy – although this is aimed mostly at academics. JSTOR is a 'not-for-profit', but it does seek a charge to cover its costs and it is mostly used in universities, public libraries and schools. If you work in the UK central government or local authority, your department **may well subscribe** and 'walk-in' users can access over 2,000 academic journals.
- If you still have close links to your old UK university, don't forget that alumnae often have remote access to their former university library databases, such as **Oxford University's Alumni** resource.

Other systematic reviews beyond the What Works centres

In the previous section we recommend using systematic reviews, not just single studies. Although expensive to commission these reviews from scratch, there are many completed reviews that are relatively easy to access.

General social policy

The [EPPI-Centre](#) at University College London mentioned above has an 'evidence library' of systematic reviews that anybody can access, although the range is not comprehensive as it only covers those studies that the Centre has been involved in.

A more comprehensive website for lists of systematic reviews in social policy is the [Campbell Collaboration Library](#). It covers education, crime and justice, social welfare and international development. Managed by a Secretariat, Campbell is organised into Coordinating Groups who manage a rigorous peer review process for high-quality systematic reviews. The reviews can be on highly topical subjects that are popular with decision-makers, such as [its reviews of 'mindfulness' in stress reduction](#) that has had 10,000 downloads.

Education

Another alternative free resource to the Education Endowment Foundation What Works Centre is the [Evidence for Impact](#) website that provides a simple evidence rating system for programmes relating to schools, along with concise evidence summaries. It has a comprehensive database of programmes available in the UK, including details on their effectiveness and cost, together with links to the providers and experts who can offer further support. The database can be easily searched by key stage, subject area and targeted group, so that results are tailored to the specific needs of your class or school.

Children's services

Dartington Social Research Unit has a one-stop-shop for children's service commissioners called [Investing in Children](#). It brings together evidence on 'what works' with the economic costs and benefits of 100 different competing interventions in children's services.

International development

For international development there is the [3ie database](#) covering both primary studies and systematic reviews of the effectiveness of social and economic interventions in low- and middle-income countries.

Rather than go through each of the above portals or websites, you may consider going straight to Google Scholar. As a tip, to find systematic reviews use the search terms 'subject' AND 'systematic review' e.g. 'mindfulness' AND 'systematic review'. If you want reviews from specific organisations you can expand the search string to specify that, e.g. 'mindfulness' AND 'systematic review' AND 'Campbell' or another organisation that produces reviews.

Last stop is a one-stop-shop

There are multiple one-stop-shops that collate evidence resources around a particular topic, but the danger is that they fall prey to '[Portal Proliferation Syndrome](#)'. In other words, producers of research can't resist the temptation to set up another hub, portal, platform or gateway to try and package information all in one place. New databases of databases may only confuse users.

Still, pooling information into a ready-to-use form for social policy professions is surely a noble cause. And bringing together databases in one place can help reduce confusion if done well. There are some great synthesising digital resources out there. The [Social Policy and Practice \(SPP\)](#) one-stop-shop for research in health and social policy responded to the proliferation of information resources by splicing together four previous national research databases relevant to policy: ChildData, AgeInfo, Planex and Social Care Online.

SPP is useful for any professional working in the field of social care or social work who can't get easy access to a university library. As well as helping to access 600 UK and international journals, it also includes the 'grey literature' of government or think-tank reports and guidance that will help those working in social care. You can find a brief guide to these UK research databases on the Alliance for Useful Evidence website.¹⁰⁶

Short, snappy bitesize summaries

Some Campbell Collaboration reviews can be up to one hundred pages long, but a time-poor decision-maker will need something shorter. Abstracts may tell you a fair amount, and there are plenty of **academic journals** that provide their abstracts – if not full articles – for free. Campbell reviews contain a short 'plain language summary', which are currently being made more accessible.

There are other readable and non-academic narratives that summarise reviews of research, written in a way that is practitioner-friendly (i.e. dropping academic jargon unless essential), such as the 2011 Scottish Government's **summaries of evidence**¹⁰⁷ on what works in stopping people from reoffending, followed by the 2013 UK Government's **summary of evidence**¹⁰⁸ on the same topic. Or the topical articles by academics on The Conversation website, including the **Hard Evidence** section.

Parliamentary **POSTNotes** are also highly digestible short PDF research summaries that anybody can access, not just MPs. Although they mostly cover science and technology, they have a growing number in social science and social policy. You should also subscribe to the Economic and Social Research Council regular **Evidence Briefings**; and if your interest is in children's policy, you can still find **Evidence Nuggets** from the What Works Centre for Children (now closed).¹⁰⁹

Consultancy and helpdesk-type services – paying the price for evidence

However, some research one-stop-shops are not open to everybody or may require a subscription or payment. But if the product is good, isn't it worth the investment? Some of us may be lucky enough to have access to university libraries (and those unsung heroes of evidence, good librarians) so that you can get behind the paywalls of prestigious research journals. But for those of us that cannot (including the authors of this guide, despite working for a large UK charity), is it worth paying somebody to track down your evidence?

One example of a fee-charging provider is the **Idox Information Service**. It has been around for 30 years now and claims to have over 200,000 users. As well as having a database of abstracts, there are consultants on hand to help you. It even has a good old fashioned library service – you can rent books out by post. Idox claim to be the 'most complete and accessible source of evidence and research for the public sector in the UK', but you have to pay for it.

There are other research consultancies that can help with locating evidence – for a list of these who are also members of the Alliance see the **website**.

One method that might be helpful is a rapid evidence assessment (REA). They aim to be fairly comprehensive, but have the downside of not being exhaustive, with a consequent risk of bias.¹¹⁰

This table from the National Foundation for Educational Research shows what you might get if you commissioned an REA, or other type of quicker review such as a scoping study:

Table D.2:

| Product | Why would I use it? | What will it give me? | How long will it take? | Suggested report length |
|----------------------------------------|----------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|-------------------------|
| Rapid review | To get a quick overview of an area of policy or practice | An overview of what we know about a specific topic. It can be based on research literature, policy or practice evidence. Literature is located through limited searching | 1-2 months | 10-page report |
| Scoping review | To find out if there is enough evidence to merit a rapid evidence assessment or full systematic review | A map of how much evidence there is on a topic | 1-2 months | 10-page report |
| Rapid evidence assessment (REA) | To get a thorough evidence synthesis to inform policy or practice. To answer questions about what is effective | A robust synthesis of the evidence, with key messages and recommendations. Literature is found through comprehensive searching | 3-5 months | 15-page report |
| Meta-analytical review | To pool results from a series of high-quality quantitative evaluations | A numeric estimate of the overall effect of the intervention being evaluated | 3-5 months | 15-page report |
| Full systematic review | To get a 'gold standard' evidence synthesis of all available evidence | A highly robust and systematic synthesis of the evidence, with key messages and recommendations. Literature is found through highly comprehensive searching | 6-12 months | 60-page report |

Source: The National Foundation for Educational Research, *Investigating what works: through reviews*

We have to tread carefully here. Endorsing commercial providers such as Idox (or potentially libelling them if we are critical) is difficult territory. But we felt we couldn't avoid mentioning evidence aggregators that demand a fee of some sort. And the reality is that one of the biggest conduits for research – academic journals – are not (yet) open access and require a fee. If we have missed any database resources that you think are useful, **let us know** so we can add them to our online resources.

Expert advice

Sometimes, passive databases are not enough. We might not even be sure exactly what our question is – and therefore what research we should be looking for. This is when a conversation with an expert may be helpful, so it is reassuring to hear of a new service being developed in Scotland, the **Evidence Bank**, developed by the Centre for Research on Families and Relationships (CRFR). As well as providing evidence, it has also piloted a special request service. The model was developed in partnership with organisations and services in the third and public sectors, and covers children, families, relationships and disability. The Evidence Bank is currently being developed as part of **What Works Scotland** and should be online soon.

In Westminster and Whitehall, academics have also started an **Evidence Information Service**¹¹¹ but it's in its infancy so watch this space to see when it is fully up and running.

If you are in a major government department in Whitehall, Scotland, Wales or Northern Ireland, you are lucky to have in-house analysts, researchers, statisticians and economists to call upon to help you find – and interpret – evidence. For instance, there is the **Government Social Research Service**.

This chapter has outlined a variety of easily accessible UK research resources for social policy and practice. However, once you have gathered your evidence, you don't want it gathering dust. We need to think about how to communicate and then act on it. How best to do that is the focus of the next and final section.

KEY MESSAGES FROM SECTION D

- Beware of **haphazard online searches** – use trusted repositories of research.
- Take advantage of the **freely available, accessible and actionable summaries of research**, such as those on the websites of nine UK What Works Centres, NHS Evidence, EPPI Centre and Campbell and Cochrane Collaborations.
- Read some of the **free, bitesize digests of research** such as the UK Parliament's POSTNotes or Economic and Social Research Council regular Evidence Briefings.
- If you can't find what you are looking for, you should **consider commissioning a consultancy or helpdesk-type service** to do a research summary, such as a rapid evidence assessment, scoping review or systematic review.
- If you are in a government department or local authority, **use your in-house analysts** to do the search for you.

SECTION E

How should you communicate your findings?

This section focuses on how to get your message across once you have the evidence you need.

There's not much point in gathering a strong body of research if it falls on deaf ears. The whole point is to transfer it into practice and change attitudes, beliefs and behaviour.

Successful communication means really putting yourself in the shoes of your audience, with many official research uptake guides saying ideally this should start with the evidence gathering process itself.¹¹²

The best bet is to avoid passive dissemination, pushing information out through the door in the hope that some of it sticks. Instead, you need to actively engage with your audiences.

For instance, the Medical Research Council's (MRC) research guide¹¹³ has some helpful advice on research uptake and features guidance based on published research (rather than expert opinion). It covers social policy as well as traditional medicine, and includes case studies in areas such as crime prevention and public health.

SIX TIPS FROM THE MEDICAL RESEARCH COUNCIL'S GUIDE ON GETTING YOUR EVIDENCE HEARD

1. **Involve stakeholders** in the choice of question and design of the research to ensure relevance.
2. Provide evidence in an **integrated and graded way**: have reviews, not individual studies, and variable length summaries that allow for rapid scanning.
3. **Take account of context**, and identify the elements relevant to decision-making, such as benefits, harms and costs.
4. **Make recommendations as specific** as possible.
5. Use a **multifaceted approach involving a mixture of interactive** – rather than didactic – educational meetings, feedback, reminders and local consensus processes.
6. Successful implementation depends on changing behaviour – often of a wide range of people. This requires a **scientific understanding of the behaviours** that need to change, the factors maintaining current behaviour and barriers and facilitators to change, and the expertise to develop strategies to achieve change based on this understanding. Further research may be needed to assist the process of implementation, and implementation research teams should include a behavioural scientist.

Source: Medical Research Council (2013) *Developing and evaluating complex interventions: new guidance*

Unfortunately, as the MRC guide alludes to, the evidence base for effective implementation remains limited. There's plenty of 'good practice guides' and advice, but most of it does not seem to be based on evidence – just what counts as expert opinion. And what evidence we do have appears to be inconsistent with a lot of current practice.

For instance, the obsession with short summaries for policymakers is based on shaky evidential grounds. A multi-arm RCT by the International Initiative for Impact Evaluation (3ie) and others¹¹⁴ explored the effectiveness of a policy brief for influencing readers' beliefs and prompting them to act.

"The results are striking, and not that reassuring for those, including 3ie, who place importance on policy briefs," then-Executive Director of 3ie, Howard White, has said.¹¹⁵ For example, the trial found that changing the nature of the brief – such as its length or including authoritative op-eds – was irrelevant to changing beliefs. However, the 'messenger' of the briefing – i.e. who presents it – did matter to whether somebody might act on it. There is a considerable amount of evidence from social psychology on the 'messenger' that we will touch on below as it gives some helpful direction on better communication of policy.

GUIDES ON RESEARCH UPTAKE

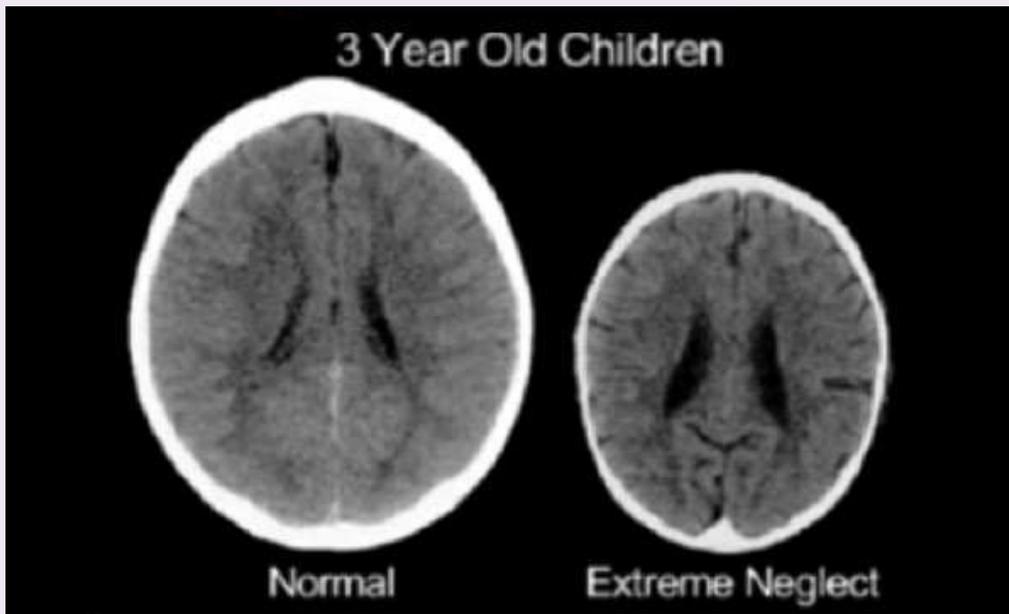
- **Research Uptake Guidance** – Department for International Development (UK)
- **Communicating research for evidence-based policymaking: A practical guide for researchers in socio-economic sciences and humanities** – European Commission
- **Guide: Engaging with Policymakers** – National Coordinating Centre for Public Engagement (UK)
- **Impact Toolkit** – Economic and Social Research Council (UK)
- **Helping researchers become policy entrepreneurs: How to develop engagement strategies for evidence-based policymaking** – Overseas Development Institute (UK)

Make it Easy, Attractive, Social and Timely

If you want to encourage a behaviour change based on research, make it **Easy, Attractive, Social and Timely (EAST)**. These four simple principles are based on the Behavioural Insights Team's own work and wider academic literature:¹¹⁶

- **Easy** – e.g. **simplify the message**. This is probably one of the most common messages of communicating research to lay audiences. This may even be as simple as the format of the written briefing. Eye-tracking research conducted for Royal Mail in 2010¹¹⁷ suggests that **people generally focus on headings, boxes and images, while detailed text is often ignored**. However, simplification of messages can be overstressed. There is some evidence that complexity has more lasting learning outcomes, if you are really trying to change people's knowledge and understanding.¹¹⁸
- **Attractive** – e.g. **use images or personalisation**. This helps to attract attention (see box on following page on use of images and MRI scans).
- **Social** – e.g. **use anecdotes, and real people**. Tap into networks to help encourage wider behaviour change, and encourage people to make a commitment to each other.
- **Timely** – we are more influenced by **immediate costs/benefits** rather than later ones (a challenge to early intervention) so make it current. Also get your timing right and prompt people when they are likely to be most receptive; the same offer made at different times can have drastically different levels of success.

THE POWER OF IMAGES – AND BRAIN SCANS



Perry, B.D. (2002) Childhood experience and the expression of genetic potential: what childhood neglect tells us about nature and nurture. 'Brain and Mind'. 3: 79-100.

This image has been highly influential and was put on the front-page of a well-evidenced and respected report for government by MP Graham Allen¹¹⁹ making the case for early intervention. The right-hand brain appears to show the impact on the brain of a child suffering 'extreme neglect' and damage for their subsequent lives, perhaps based on bad parenting. Some, however, have queried the use of such MRI images.¹²⁰ What, for instance, was the case-history of the two children's brains above? Perhaps the child on the right had other severe disabilities.

But we do know how powerful neuro-imaging can be on policymakers – and even on other experts. One RCT experiment found that including brain images was linked to higher perceptions of scientific merit. University students were shown identical academic papers, the only difference being that one included brain images, one had none (and one with other images, like traditional bar graphs). People thought the paper with the brain image had much more scientific credibility, without realising that it was exactly the same text and research.¹²¹

KEY MESSAGES FROM SECTION E

- Successful communication means **empathising with your target audience** – ideally from the start of the evidence gathering process.
- Avoid passive dissemination. Instead, you need to **actively engage with your audiences**.
- If you do have to be more passive in your communication, use the Behavioural Insights Team's tips on behaviour change. **Make it Easy, Attractive, Social and Timely**.

SUMMARY

Whether you are a policymaker or leader in a charity or local authority, evidence is unlikely to be a nice 'fit' with all the other competing sources of information. Research evidence is – and should be – just part of the mix.

How to go about smarter decision-making generally is a whole new practice guide in itself. Only pointers and checklists can be given here. There will never be a magic formula for how to 'do' evidence-informed policy. There are, however, models and tools to help decision-making. For policymakers, [Harvard's Evidence for Policy Design](#) team has created online tools such as 'decision-trees' or a 'policy analysis matrix' to help weigh up the competing pieces of information behind a policy, thanks to funding from the UK's DFID.¹²² These decision-making tools are helpful but will never replace professional judgment in the messy and complex world of social policy. They can be a prop or checklist to support it. Instead of our own list, we have two key recommendations for using evidence to inform better decision-making:

1. Implement changes or advice based on evidence that is as strong as possible (and vice versa, be wary of changes or advice if the evidence is weak)

The advice of this guide is that the stronger the evidence, particularly if it is based on not just one study but multiple replications and systematic reviews, the more you should be confident of it being a part of your decision.

What we shouldn't be doing is making bolder policy recommendations based on weaker evidence. That may seem obvious to you, but it doesn't seem to be happening in policymaking. The National Audit Office review of government evidence (see page 29) found that the weakest evaluations were more likely to make 'bold, un-caveated claims' about the positive impacts of the policy examined.¹²³ If there is weak – or non-existent – evidence, then be honest about it and say so. Over-claiming is a dangerous game. And although we know how frustrating it is to see a conclusion of 'we need more research' in policy reports or advice, it's also true: we will never stop needing to understand and research.

2. Be evidence-aware in your decision-making

The second piece of advice is to be self-aware about potential cognitive biases in your decision-making. Being alert to the fact that you may be leaping to the wrong conclusions – even when you have located the best available research – is paramount.

Finally, we want to stress that the single most important message of this guide is appropriateness: 'horses for courses'. You need to think about the right type of research to suit your needs. It has to be proportionate – a massive, multi-armed RCT would be madness for a small rural-based social enterprise. But likewise, a decade-long £50 million social programme should seriously invest in evidence.

The type of evidence also has to match your question. Does it fit where you are in growing a programme, policy or practice? To help you make that judgement, use some of the formal framework, such as the Nesta Standards of Evidence, if your question is about impact. Also, think about where you are in the Nesta innovation spiral – would learning from others be your best choice, or is it time to formally experiment on your growing idea?

And don't be put off by expense – we have given you some handy pointers to where research is already available, and for free. Research is becoming easier to source. Academics are under more pressure to step outside their university walls and help you in policymaking, charitable work or delivering good local services.

We also need to be smarter on how to communicate evidence. Evidence rarely speaks for itself. A common communication message is to simplify, visualise, empathise and be 'translational' – in other words, interact with your audience. In the coming months and years, we will be doing just that: creating videos and interactive digital media for blended learning, and perhaps, dear reader, meeting you face-to-face in one of our ongoing Evidence Masterclasses.

APPENDIX 1

Rationales for evidence and types of evidence required for policymakers

Source: Louise Shaxson (2014) Investing in Evidence; Lessons from the UK Department for Environment, Food and Rural Affairs

| Heading | 'Big questions' | Rationales for evidence needs | Types of evidence required |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A: Understanding the context ; fundamental processes and phenomena, baselines and benchmarks | Where are we now? | <ul style="list-style-type: none"> • To gather and analyse available/new data • To evaluate risks, issues and uncertainties | <ul style="list-style-type: none"> • Reviews of existing knowledge • Surveys of social and environmental data • Research on causality • Risk assessment |
| B: Development of models, methodologies and tools | Where are we going? | <ul style="list-style-type: none"> • To understand current drivers and trends • To predict future drivers and trends • To assess implications for policy outcomes | <ul style="list-style-type: none"> • Sensitivity analysis • Horizon scanning • Forecasting and scenarios • Modelling impacts and outcomes |
| C: Developing and using the evidence base to help set targets and formulate policy | Where do we want to be over the next 5-10 years? | <ul style="list-style-type: none"> • To understand the economic/social value of change • To understand the feasibility/cost of change • To negotiate goals | <ul style="list-style-type: none"> • Economic and social research • Deliberative engagement processes • Feasibility and pilot studies • Market surveys |
| D: Development and appraisal of options/solutions E: Optimum decisions and effective implementation through communication , engagement and consultation to influence change | How do we get there? | <ul style="list-style-type: none"> • To identify/evaluate current options • To identify/develop new solutions • To evaluate new/old options | <ul style="list-style-type: none"> • Option/evaluation studies • Regulatory impact assessments • Interventions to promote innovation |
| F: Monitoring progress towards policy/programme targets G: Policy/programme evaluation | How well did we do? | <ul style="list-style-type: none"> • To monitor progress • To evaluate policies and programmes • To learn lessons | <ul style="list-style-type: none"> • Interdisciplinary evaluations • Deliberative evaluation processes |

APPENDIX 2

Experimental designs for evaluating complex interventions

This list of five different experimental designs is taken from *Developing and evaluating complex interventions: new guidance*, prepared on behalf of the Medical Research Council (2013).

1. Individually randomised trials

Individuals are randomly allocated to receive either an experimental intervention, or an alternative such as standard treatment, a placebo or remaining on a waiting list. Such trials are sometimes dismissed as inapplicable to complex interventions, but there are many variants of the basic method, and often solutions can be found to the technical and ethical problems associated with randomisation.

2. Cluster randomised trials

Contamination of the control group, leading to biased estimates of effect size, is often cited as a drawback of randomised trials of population level interventions, but cluster randomisation, widely used in health service research, is one solution. Here, groups such as patients in a GP practice or tenants in a housing scheme are randomly allocated to the experimental or a control intervention.

3. Stepped wedge designs

The randomised stepped wedge design may be used to overcome practical or ethical objections to experimentally evaluating an intervention for which there is some evidence of effectiveness, or which cannot be made available to the whole population at once. It allows a randomised controlled trial to be conducted without delaying roll-out of the intervention. Eventually, the whole population receives the intervention, but with randomisation built into the phasing of implementation.

4. Preference trials and randomised consent designs

Practical or ethical obstacles to randomisation can sometimes be overcome by the use of non-standard designs. Where patients have very strong preferences among treatments, basing treatment allocation on patients' preferences, or randomising patients before seeking consent, may be appropriate.

5. N-of-1 designs

Conventional trials aim to estimate the average effect of an intervention on a population, and provide little information about person variability in response to interventions, or about the mechanisms by which effective interventions achieve change. N-of-1 trials, in which individuals undergo interventions with the order or scheduling decided at random, can be used to assess between and within person change, and to investigate theoretically predicted mediators of that change.

APPENDIX 3

Guidance on evidence quality, designs and methods

Scottish Government Social Researchers' Method Guides – short introductions to some of the most common methods used in social science research

www.gov.scot/Topics/Research/About/Social-Research/Methods-Guides

The Bond Evidence Principles and checklist – specially designed for NGOs

www.bond.org.uk/effectiveness/monitoring-and-evaluation

Evidence for Success: the guide to getting evidence and using it – a practical guide for third sector organisations from Evaluation Support Scotland and KTN

www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised_april_2015-web.pdf

Research Guide for Third Sector Organisations

www.ncb.org.uk/media/858179/research_guide_for_third_sector_organisations.pdf

The Medical Research Council's Developing and evaluating complex interventions: new guidance – provides guidance on the development, evaluation and implementation of complex interventions to improve health

www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/

Government Office for Science – The Government Chief Scientific Adviser's Guidelines on the Use of Scientific and Engineering Advice in Policymaking

www.gov.uk/government/uploads/system/uploads/attachment_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf

National Institute for Health and Care (NICE) – The Guidelines Manual: Reviewing the Evidence

www.nice.org.uk/article/pmg6/chapter/6-reviewing-the-evidence

DFID – How to Note: Assessing the Strength of Evidence

www.gov.uk/government/publications/how-to-note-assessing-the-strength-of-evidence

HM Treasury – The Magenta Book: Guidance for Evaluation

www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

Civil Service Learning – Policy Profession: Skills and Knowledge Framework

civilservicelearning.civilservice.gov.uk/sites/default/files/policy_profession_skills_and_knowledge_framework_jan2013web.pdf

HM Treasury, DECC and DEFRA – Quality in policy impact evaluation; understanding the effects of policy from other influences (Supplement to Magenta Guide)

www.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation_QPIE_.pdf

Research Councils, Universities UK *et al.* – UK concordat to support research integrity

www.universitiesuk.ac.uk/highereducation/Pages/Theconcordattosupportresearchintegrity.aspx#.Vd2X5flVhBc

Guides to commissioning research

Social Research Association – Commissioning Social Research good practice guide

the-sra.org.uk/sra_resources/research-commissioning/

Market Research Society guide to ‘buying research’

www.mrs.org.uk/intelligence/research_buyers_guide

ENDNOTES

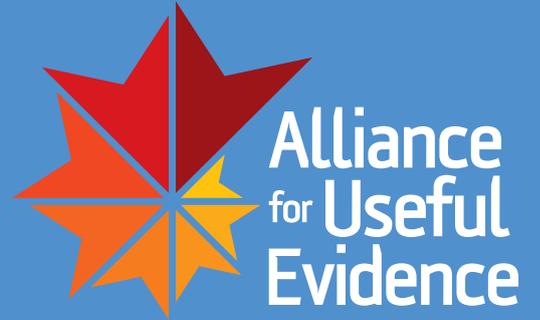
1. The Police Foundation (2013) Policing and Crime Reduction: The evidence and its implications for practice. See page 11: http://www.police-foundation.org.uk/uploads/holding/projects/policing_and_crime_reduction.pdf
2. Sense about Science (2015) Making Sense of Crime: Evidence on the causes of crime and policies to reduce it. See page 17.
3. OECD (2015) Students, Computers and Learning: Making the Connection. See: <http://www.oecd.org/publications/students-computers-and-learning-9789264239555-en.htm>
4. For an overview of evidence on 'what works' in the classroom for pupil attainment, see the Education Endowment Foundation Toolkit: <https://educationendowmentfoundation.org.uk/toolkit/toolkit-a-z/>
5. This is taken from the 'Policy Profession Skills and Knowledge Framework' (see endnote 9) but it is just as relevant to our audience of this guide – working outside policy, in local authorities and charities.
6. Sackett, D., Rosenberg, W., Gray, M., Haynes, B. and Richardson, S. (1996) Evidence based medicine: what it is and what it isn't. 'British Medical Journal'. 312:71. See: <http://www.bmj.com/content/312/7023/71>
7. The definition came at a time when most medical decision-making was based on experience, authority and eminence. Medical practice was not informed by the best available scientific evidence. Commentators and researchers like Ben Goldacre have argued that social policy is in the same place as medicine 20 years ago: authority still reigns, not research.
8. Graber, M.L., Franklin, N. and Gordon, R. (2005) Diagnostic Error in Internal Medicine. 'Internal Medicine.' 165(13). 1493-1499. See: <http://archinte.jamanetwork.com/article.aspx?articleid=486642&resultclick=1>
9. Civil Service Learning (2013) 'Policy Profession Skills and Knowledge Framework.' London: Civil Service Learning. See: https://civilservicelearning.civilservice.gov.uk/sites/default/files/policy_profession_skills_and_knowledge_framework_jan2013web.pdf
10. <http://www.oxforddictionaries.com/definition/english/evidence>
11. Frost, S. *et al.* (2006) 'The Evidence Guide; Using Research and Evaluation in Social Care and Allied Professions.' London: Barnardo's
12. Nutley, S., Powell, A. and Davies, H. (2013) 'What Counts as Good Evidence?' London: Alliance for Useful Evidence
13. Adapted from Shaw, I. (1999) 'Qualitative Evaluation.' London: Sage.
14. Tetlock, P.E. (2005) Expert Political Judgement: How Good Is It? How Can We Know?. New Jersey: Princeton University Press
15. Vermeren, P. (2015) 'Why the CEBMa model is misleading.' See: http://evidencebasedhrm.be/wp-content/uploads/2015/08/Why-the-Cebma-model-is-misleading_august_2015.pdf
16. Ariely, D. (2009) Predictably Irrational: The Hidden Forces that Shape Our Decisions. London: HarperCollins.
17. See also White, H. and Phillip, D. (2012) 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework.' – a working paper from the International Initiative for Impact Evaluation that discusses how evaluators themselves can be very biased: "Bias affects all evaluation approaches and methods; it affects the choices that evaluators make, the questions that they ask, the data they collect and the way that they carry out analysis" (page 24).
18. The chances of social workers getting it wrong are compounded by the sheer burden on their decision-making – they have to make many tough sequential decisions through the course of a single day, which engenders 'decision fatigue'. For a discussion of this research see page 7 of Breckon, J. and Hay, J. (2015) 'Knowing How to Protect; Using Research Evidence to Protect Children from Harm'. London: Alliance for Useful Evidence/NSPCC. See: <http://www.alliance4usefulevidence.org/assets/NSPCC-report-FINAL-V-for-printers.pdf>
19. Kassina, S.M., Drorb, I.E. and Kukuckaa, J. (2013) The forensic confirmation bias: Problems, perspectives, and proposed solutions. 'Journal of Applied Research in Memory and Cognition.' 2(1). 42-52. See: http://cognitiveconsultantsinternational.com/Dror_JARMAC_forensic_bias.pdf
20. Kahneman, D., Lovallo, D. and Sibony, O. (2012) 'The Big Idea: Before You Make That Big Decision...'. See: <https://hbr.org/2011/06/the-big-idea-before-you-make-that-big-decision>
21. Flyvbjerg, B., Skamris Holm, M.K. and Buhl, S.L. (2002) Underestimating Costs in Public Works Projects – Error or Lie. 'Journal of the American Planning Association.' 68(3). 279-295.
22. See page 7 of Lewis, P.G. (2013) 'Policy Thinking, Fast and Slow: A Social Intuitionist Perspective on Public Policy Processes'. APSA 2013 Annual Meeting Paper, American Political Science Association 2013 Annual Meeting. See: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2300479
23. See Cairney, P. (2014) 'The Psychology of Evidence Based Policymaking: Who Will Speak For the Evidence if it Doesn't Speak for Itself?' See: <https://paulcairney.wordpress.com/2014/04/07/1288/>
24. Kahneman D. and Klein G. (2009) Conditions for intuitive expertise: a failure to disagree. 'American Psychology.' 64(6). 515-26. See: <http://www.ncbi.nlm.nih.gov/pubmed/19739881>
25. Gornall, J. (2008) Does cot death still exist? 'British Medical Journal.' 336(7639). 302-304. See: <http://dx.doi.org/10.1136/bmj.39455.496146.AD>
26. Nesta (2013) 'Understand how innovation works.' Video available at <http://www.nesta.org.uk/resources/understand-how-innovation-works>

27. Gabriel, M. (2014) 'Making it Big: Strategies for scaling social innovations.' London: Nesta. See: <http://www.nesta.org.uk/publications/making-it-big-strategies-scaling-social-innovations>
28. White, H. (2015) 'Learning what works and why from evidence of what doesn't work.' See: <http://www.alliance4usefulevidence.org/learning-what-works-and-why-from-evidence-of-what-doesnt-work/>
29. Nesta and TSIP (2014) 'Guidance for Developing a Theory of Change for Your Programme.' See: https://www.nesta.org.uk/sites/default/files/theory_of_change_guidance_for_applicants_.pdf
30. See page 3 of Stern, E. (2015) 'Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.' London: Big Lottery Fund, Bond, Comic Relief and the Department for International Development.
31. Johnson, S. (2011) 'Impact of social science on policy: Sure Start case study.' Report to Economic and Social Research Council (ESRC)
32. Mulgan, G. (2015) 'The six Ws: a formula for what works.' London: Nesta. See: <http://www.nesta.org.uk/blog/six-ws-formula-what-works>
33. Nesta (2013) 'What is ethnographic research?' Video available at <https://vimeo.com/77509484>
34. Petticrew, M. and Roberts, H. (2003) Evidence, hierarchies and typologies: horses for courses. 'Journal of Epidemiology and Community Health.' 57. 527-529.
35. For an authoritative and exhaustive list of social science research frameworks and methods see Luff, R., Byatt, D. and Martin, D. (2015) 'Review of the Typology of Research Methods within the Social Sciences.' London: ESRC/National Centre for Research Methods. This typology is one of the most downloaded parts of the NCRM website. See: <http://eprints.ncrm.ac.uk/3721/>
36. A 2015 recent guide to impact evaluation made the same point. The paper by Elliot Stern, commissioned by the Big Lottery Fund, Bond, Comic Relief and the Department for International Development, said that one of the problems faced by those approaching impact evaluations is that: "it is often presented as a technical or methodological question only accessible to experts or researchers. To some extent this is true but the main arguments, logics and choice-points are more accessible."
37. Department for International Development (2014) 'How to Note: Assessing the Strength of Evidence.' See: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf
38. For more on the importance of anecdote, see Greenhalgh, T. (2014) 'Stories or numbers OR stories and numbers.' See: <http://www.alliance4usefulevidence.org/stories-or-numbers-or-stories-and-numbers/>
39. For a fundamental critique see Frideres, J. S. (1992) 'Participatory research: An illusionary perspective.' in Frideres, J.S. (ed.) A world of communities: Participatory research perspectives. North York, Ontario: Captus University Publications. Frideres has, of course, been heavily criticised by the participatory research community.
40. This is adopted from Table 2, page 18 of Stern, E. (2015) 'Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.'
41. White, H. and Phillips, D. (2012), 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework.' See: http://www.3ieimpact.org/media/filer_public/2012/06/29/working_paper_15.pdf
42. <http://www.nesta.org.uk/publications/better-public-services-through-experimental-government>
43. Blair, P.S., Fleming, P.J., Smith, I.J., Platt, M.W., Young, J., Nadin, P. *et al.* (1999) Babies sleeping with parents: case-control study of factors influencing the risk of the sudden infant death syndrome. 'British Medical Journal.' 319 (7223). 1457-1462.
44. Medical Research Council (2013) 'Developing and evaluating complex interventions: new guidance.' See: <http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>
45. See: <http://www.alliance4usefulevidence.org/publication/spark-guide-to-testing-social-policy-innovation/>
46. See page 5 of Stern, E. (2015) 'Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.'
47. See page 21, *ibid.*
48. Carter, R. (2012) 'Helpdesk Research Report: Theory-based evaluation approach.' Birmingham: Governance and Social Development Resource Centre, University of Birmingham. See: <http://www.gsdr.org/docs/open/hdq872.pdf>
49. Westthorp, G. (2014) 'Realist Evaluation: An Introduction.' London: Methods Lab, Overseas Development Institute. See: <http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf>
50. Mayne, J. (2012) Contribution Analysis: Coming of Age?. 'Evaluation.' 18(3). 270-280.
51. For network analysis and process tracing see: Befani, B. and Mayne, J. (2014) Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation. 'IDS Bulletin.' 45(6). 17-36. See: <http://onlinelibrary.wiley.com/doi/10.1111/1759-5436.12110/abstract>
52. Baptist, C. and Befani, B. (2015) 'Coffey How To Guide: Qualitative Comparative Analysis - A Rigorous Qualitative Method for Assessing Impact.' See: <http://www.coffey.com/assets/Ingenuity/Qualitative-Comparative-Analysis-June-2015.pdf>
53. See the Scottish Government Social Researchers' Method Guide on Contribution Analysis: <http://www.gov.scot/resource/doc/175356/0116687.pdf>
54. For example the old guide by the late, great evaluation guru Carol Weiss: Weiss, C. (1998) Evaluation: methods for studying programs and policies. New York: Prentice Hall.
55. <http://www.thinknpc.org/blog/getting-started-with-cost-benefit-analysis/>
56. Cabinet Office (2007) 'A guide to Social Return on Investment.' See: https://www.bond.org.uk/data/files/Cabinet_office_A_guide_to_Social_Return_on_Investment.pdf

57. HM Treasury (2011) 'The Green Book: Appraisal and Evaluation in Central Government.' London: HM Treasury.
58. <https://www.nao.org.uk/report/choosing-the-right-fabric-3/>
59. <https://www.gov.uk/government/publications/supporting-public-service-transformation-cost-benefit-analysis-guidance-for-local-partnerships>
60. Petticrew, M. and Roberts, H. (2003) Evidence, hierarchies and typologies: horses for courses. 'Journal of Epidemiology and Community Health.' 57. 527-529.
61. Ioannidis, J.A. (2005) Contraindications and Initially Stronger Effects in Highly Cited Clinical Research. 'The Journal of the American Medical Association.' 294(2). 218-228.
62. Cited in The Economist (2013) 'Unreliable research; trouble in the lab.' 409(8858). 24.
63. Easterbrook, P.J., Gopalan, R., Berlin, J.A. and Matthews, D.R. (1991) Publication bias in clinical research. 'The Lancet.' 337(8746). 867-872.
64. Social sciences suffer from severe publication bias; the survey found that 'null results' rarely see the light of the day.
65. Wolford, B. (2014) 'Big Pharma Plays Hide-the-Ball With Data.' See: <http://europe.newsweek.com/medical-science-has-data-problem-284066?rm=eu>
66. Sense about Science (2006) 'I don't know what to believe: Making sense of science stories.' See: <http://www.senseaboutscience.org/resources.php/16/i-dont-know-what-to-believe>
67. See page 10 of Government Office for Science (2010) 'The Government Chief Scientific Adviser's Guidelines on the Use of Scientific and Engineering Advice in Policy Making.' See: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf
68. See page 14 of http://www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised_april_2015-web.pdf
69. National Audit Office (2013) 'Evaluation in Government.' See: <http://www.nao.org.uk/report/evaluation-government/#>
70. See for instance, see this guide as a supplement to the Magenta Guide: HM Treasury, DECC and DEFRA (2012) 'Quality in policy impact evaluation; understanding the effects of policy from other influences.' The guidance shows how 'higher quality research designs can help meet the challenge of attributing measured outcomes to the policy in question (as opposed to other influences), whereas lower quality designs reduce confidence in whether it was the policy that achieved those outcomes' (page 5).
71. For a list and short description of these scales and hierarchies, see Annex 1 of Nutley, S., Powell, A. and Davies, H. (2013) 'What Counts as Good Evidence?.' London: Alliance for Useful Evidence.
72. For a model of different approaches on quality that includes four dimensions: (1) methodological quality; (2) quality in reporting; (3) appropriateness; (4) relevance to policy and practise see: Boaz, A. and Ashby, D. (2003) 'Fit for purpose? Assessing research quality for evidence based policy and practice.' ESRC UK Centre for Evidence Based Policy and Practice Working Paper 11.
73. See the Research Councils UK and Universities 'UK concordat to support research integrity', signed in 2012 by many key research funders: <http://www.universitiesuk.ac.uk/highereducation/Pages/Theconcordattosupportresearchintegrity.aspx#.Vd2X5flVhBc>
74. Callaway, E. (2011) Report finds massive fraud at Dutch universities. 'Nature.' 479(15). See: <http://www.nature.com/news/2011/111101/full/479015a.html>
75. Boaz, A. and Ashby, D. (2003) 'Fit for purpose? Assessing research quality for evidence based policy and practice.' ESRC UK Centre for Evidence Based Policy and Practice Working Paper 11.
76. HM Treasury, DECC and DEFRA (2012) 'Quality in policy impact evaluation; understanding the effects of policy from other influences.'
77. HM Treasury (2011) 'The Magenta Book: Guidance for Evaluation.' London: HM Treasury.
78. HM Treasury (2011) 'The Green Book: Appraisal and Evaluation in Central Government.' London: HM Treasury.
79. Brooks, G., Hall, J. and Torgerson, C.J. (2006) 'A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling.' London: Department for Education and Skills.
80. Henderson, M. (2013) 'The Geek Manifesto: Why science matters.' London: Corgi. Page 17.
81. HM Treasury (2011) 'The Magenta Book: Guidance for Evaluation.' London: HM Treasury.
82. See Morgan, S.L. (ed.) (2013) 'Handbook of Causal Analysis for Social Research.' Berlin: Springer. But it has to be confronted head-on: <http://www.tylervigen.com/spurious-correlations>
83. See page 4 of Stern, E. (2015) 'Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.'
84. Qualitative research explores and tries to understand people's beliefs, experiences, attitudes, behaviour and interactions. It generates non-numerical data. Well-known qualitative research methods include in-depth interviews, focus groups, documentary analysis and participant observation, for example.
85. Although over a decade old, the Cabinet Office and defunct Office for the Chief Social Researcher have a guide on the quality of qualitative research. Regardless of where you stand in the paradigm wars, there's enough consensus that quality matters – and that we can't avoid making judgements about what is good or not. See: <http://www.cebma.org/wp-content/uploads/Spencer-Quality-in-qualitative-evaluation.pdf>
86. For an overview – and critique – of hierarchies of evidence, see: <http://www.alliance4usefulevidence.org/assets/What-Counts-as-Good-Evidence-WEB.pdf>
87. Public Health England (2014) 'Identifying what works for local physical inactivity interventions.'
88. https://www.nesta.org.uk/sites/default/files/nesta_standards_of_evidence_csaif.pdf
89. The Nesta and Project Oracle standards were strongly inspired by the independent Social Research Unit at Dartington, who work with many local authorities and children's service providers.

90. Ní Ógáin, E. (2015) 'Impact measurement in impact investing learning from practice.' London: Nesta Impact Investments. See: http://www.nesta.org.uk/sites/default/files/impact_measurement_in_impact_investing.pdf
91. Ibid
92. Ruiz, F. and Breckon, J. (2014) 'The NICE way; lessons for social policy and practice from the National Institute for Health and Care Excellence.' London: Alliance for Useful Evidence/NICE.
93. Rawlins, M. (2014) Evidence, values and decision-making. 'International Journal of Technology Assessment in Health Care'. 30(2). 233-8.
94. See for example page 28 of Nutley, S. *et al.* (2013) 'What Counts as Good Evidence?'. London: Alliance for Useful Evidence.
95. To see the difference between systematic reviews and meta-analysis, see pages 64-5 of 'The Magenta Book.' For a quick guide on doing reviews, aimed at policymakers see Alliance for useful Evidence (2013) 'Learning from Research: Systematic Reviews for Informing Policy Decisions.': <http://www.alliance4usefulevidence.org/assets/Alliance-FUE-reviews-booklet-3.pdf>
96. For a critique of applying research syntheses into policy, see Pawson, R. (2001) 'Evidence Based Policy: In Search of a Method.' ESRC UK Centre for Evidence Based Policy and Practice: Working Paper 3.
97. See page 9 of <http://www.alliance4usefulevidence.org/assets/Alliance-final-report-08141.pdf>
98. HM Treasury (2011) 'The Magenta Book: Guidance for Evaluation.' London: HM Treasury.
99. Thomas, J., Newman, M., Oliver, S. (2013) Rapid evidence assessments of research to inform social policy: taking stock and moving forward. 'Evidence & Policy: A Journal of Research, Debate and Practice.' 9(1). 5-27.
100. <http://www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/dfe-evidence-check-forum/>
101. Begley, C.G. and Ellis, L.M. (2012) Drug development: Raise standards for preclinical cancer research. 'Nature.' 483. 531-533.
102. <http://www.alliance4usefulevidence.org/replication/>
103. McCormick, J. (2013) 'Evidence Exchange: Learning from social policy from across the UK.' London: Carnegie UK Trust and Joseph Rowntree Foundation.
104. <http://www.nesta.org.uk/blog/why-what-works-centres-are-working>
105. https://www.nesta.org.uk/sites/default/files/nesta_a4ue_and_nice_paper.pdf
106. <http://www.alliance4usefulevidence.org/a-brief-guide-to-uk-social-policy-databases/>
107. <http://www.scotland.gov.uk/Resource/0038/00385880.pdf>
108. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/243718/evidence-reduce-reoffending.pdf
109. Although this guide is for UK social policy, it's also worth mentioning 3ie short guides in international development. See: <http://www.3ieimpact.org/en/evidence/policy-briefs/3ie-briefs/>
110. <http://www.alliance4usefulevidence.org/rapid-evidence-assessments-a-bright-idea-or-a-false-dawn/>
111. <http://www.alliance4usefulevidence.org/the-creation-of-a-new-service-to-unlock-research-expertise-and-you-hold-the-key/>
112. E.g. European Commission (2010) 'Communicating research for evidence-based policymaking - a practical guide for researchers in socio-economic sciences and humanities.' See: <http://www.spia-europa.de/pdf/guide-communicating-research.pdf>
113. Medical Research Council (2013) 'Developing and evaluating complex interventions: new guidance.' See: <http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>
114. Beynon, P. *et al.* (2012) 'What Difference does a Policy Brief Make?.' International Initiative for Impact Evaluation (3ie), Institute of Development Studies, Norwegian Agency for Development Cooperation. See: http://www.3ieimpact.org/media/filer_public/2012/08/22/fullreport_what_difference_does_a_policy_brief_make__2pdf_-_adobe_acrobat_pro.pdf
115. http://www.3ieimpact.org/media/filer_public/2012/08/24/3ie_policy_brief_experiment_summary.pdf
116. Behavioural Insights Team (2014) 'EAST: Four Simple Ways to Apply Behavioural Insights.' See: http://www.behaviouralinsights.co.uk/wp-content/uploads/2015/07/BIT-Publication-EAST_FA_WEB.pdf
117. Research quoted in Royal Mail (undated) 'Door to Door; Big Return. The seven secrets of highly effective leaflet design.' See: http://www.royalmail.com/sites/default/files/docs/pdf/seven_secrets_from_rm_v1.09.pdf
118. See Rousseau, D.M. (ed.) (2012) Oxford Handbook of Evidence Based Management. Oxford: Oxford University Press.
119. Allen, G. (2011) 'Early Intervention: The Next Steps An Independent Report to Her Majesty's Government.' See: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/284086/early-intervention-next-steps2.pdf
120. <http://www.theguardian.com/society/2014/may/06/policymakers-neuroscience-justify-early-intervention-agenda-bruce-perry>
121. McCabe, D.P. and Castel, A.D. (2007) Seeing is believing: The effect of brain images on judgments of scientific reasoning. 'Cognition.' 107 (1). 343-352. See: <http://www.imed.jussieu.fr/en/enseignement/Dossier%20articles/article4.pdf>
122. This is part of the Building Capacity to Use Research Evidence (BCURE) programme of work, funded by the UK Department for International Development.
123. <http://www.alliance4usefulevidence.org/the-poor-state-of-evaluation-in-government/>

Nesta...



Nesta/Alliance for Useful Evidence

1 Plough Place
London EC4A 1DE

www.alliance4usefulevidence.org

www.nesta.org.uk

