Evaluating the Impact of Research Programmes - Approaches and Methods

Contents

Approaches and Methods

Many of these methods are used alongside one another in a single evaluation.
Approaches:

- Case Studies
- Episode Studies

Frameworks specifically for evaluating research impact:

- HERG Payback
- SIAMPI
- NIHR 'lean'/dashboard approach
- RAPID Outcome Assessment

Relevant methods and tools (not comprehensive):

- Contribution analysis
- Outcome Mapping
- Process tracing
- Most Significant Change
- Social Network Analysis
- Actor Network Theory
- Qualitative Document Analysis
- Surveys
- Cohort Studies
- Randomised Controlled Trials
- Alternative Indicators
- Digital footprint analysis
- SenseMaker

Issues to consider:

- Recipient-led impact
- What to choose?

Case studies

These focus on understanding a unit (person, site or project – in this case a research programme) in its context, which can use a combination of qualitative and quantitative data.

Case studies have strengths:

- They can be descriptive and explanatory, and rich in detail.
- Can demonstrate pathways from research to impact
- Potential to combine sources and methods (triangulation)
- Explore context
- Particularly useful in situations and contexts where the understanding of research impacts is poor, and there is only some knowledge on the causalities involved

And weaknesses:

- Difficulty in selecting cases
- Issues of bias and attribution
- Some experts at the workshop considered that case studies tend to overplay the role of research in causing change - 'supply perspective'
- Time-intensive to ensure rigour
- Difficult to apply a common framework across case studies. Must follow a standardised protocol in order to be comparable.
- Often poor generalisability of findings
- Highly dependent on the skill of the researcher
- Often poor generalisability of findings

Although impact findings from case studies can be difficult to generalise, lessons can be learned from a portfolio of case studies by analysing the common factors in these pathways that facilitated impact generation, ("qualitative comparative analysis") – though it can be difficult to apply a common framework.

Episode Studies

These track back from policy change and understand sequence of events leading to it. The approach uses a four element framework (political context, use of evidence,

links and relationships, external factors) to identify factors, key actors and events that have contributed to a policy change. Data collection is from mixed methods used, including interviews and documentary analysis.

Episode Studies are good at assessing the relative importance of influences, and the process of working backwards can capture complex policy processes. However, like all methods, Episode Studies have weaknesses: there is a risk of actors 're-writing' history, and obviously, isn't so useful for evaluators wanting to look forwards to capture outcomes from research programmes.

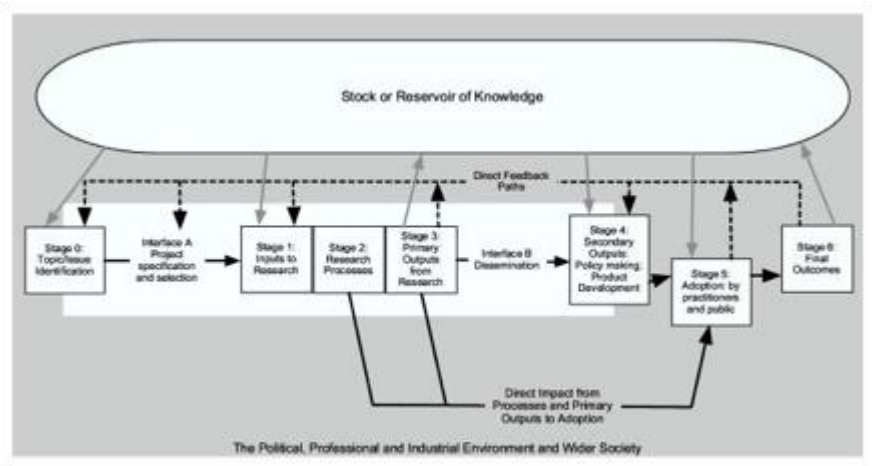
Overall, experts at the workshop concluded Episode Studies tend to emphasise the role of politics and show research as less significant in causing change (contrast with Case Studies).

See John Young's slides.

HERG Payback

The Health Economics Research Group (HERG) at Brunel University developed a framework to trace impacts from specific bodies of health research. The framework has two main elements:

1. **A multidimensional categorisation of benefits or paybacks starting with more traditional academic benefits of knowledge production and research capacity-building, and then extending to wider benefits to society:**
    - knowledge production
    - targeting future research, capacity building and absorption
    - informing policies and product development
    - health and health sector benefits, e.g. better health, health equity etc.
    - broader economic benefits
2. **A logic model of how to assess the benefits. Expected or identified benefits are mapped onto a model depicting the research process itself and the stages at which each category of benefits would be realised. The model consists of:**
    - 7 stages
    - 2 interfaces



This Payback Framework was originally commissioned by the English Department of Health in the 1990s, but has been applied to a wide range of research in many countries including Hong Kong and Bangladesh (by ICDDR,B)

See this paper for an explanation of the Framework

More recently the MRC, Wellcome Trust and Academy of Medical Sciences funded HERG and colleagues to expand their work in an attempt to evaluate the economic returns from health research to the UK, and while this approach is yet to be used in a development context there are potentially ways in which it could be adapted for this purpose.

Stephen Hanney outlined his thoughts on the strengths and weaknesses of the HERG Payback framework and its applicability in development. See Stephen Hanney's slides here.

External reviewers have also highlighted strengths and weaknesses of the model.

SIAMPI

The "Social Impact Assessment Methods through the study of Productive Interactions" (SIAMPI) approach to the assessment of research impact is characterised by a focus on processes and seeks to understand them by targeting 'productive interactions' between researchers and stakeholders and their discussions about achieving agreed social goals. These 'productive interactions' are those interactions in a network that produce value (not only economic, but also socio-cultural, ethical, intellectual, technological, environmental) and lead to behavioural change.

These could be

1. **Personal interactions : joint projects, advisory, consultancy, double functions,**

mobility

2. **Interactions through media**
   a. Texts : articles, books, catalogues, protocols, new diagnostics
   b. Artefacts : instruments, exhibitions, models, designs

3. **Support: contracts, subsidies, patenting, licensing, sharing of people and facilities**

The overall approach of this method of impact evaluation is to

- Analyse the mission of the research programme/group
- Involve the relevant context from early on [audiences/stakeholders]
- Carry out assessment (incl. indicators) as a joint effort between researchers and stakeholders
- Seek to increase commitment and awareness through focus groups

As an evaluation tool it focuses on mutual learning instead of accountability (see Motivations for evaluating impact). It helps to overcome some of the common problems with impact evaluation (attribution/contribution; temporality/small steps; robustness of indicators; commitment of participants). It has been tested in many different fields and in a few different countries but is yet to be tested in international development.

SIAMPI has a potential weakness in that there is no top down method to determine what's important – you have to trust the people doing the research and their perceptions of impact.

See here for the latest summaries and preliminary results of SIAMPI case studies.

See Jack Spaapen's slides here.

NIHR 'lean'/dashboard approach

The UK's National Institute for Health Research (NIHR) is a deliberate intervention to embed research into the NHS, with the goal of improving the health and wealth of the nation. For the evaluation team there is a clear goal (see Why and what? Motivations for evaluating impact and tools for framing the right question) to ensure that the Treasury keeps paying to support research in the NHS – and this shapes the methodology they have chosen.

David Kryl described his work laying the groundwork to measure the outcomes of research. He is collecting data to allow comparison of survival rates before and after

NIHR activity. Early indications show big differences in colon cancer.

It is easier to measure the inputs, outputs and impacts of the research infrastructure (rather than the research itself, mentioned above). His methodology uses a dashboard to track performance. An annual survey feeds into this straightforward dashboard, and helps the team find the right stories to complement the data. All the metrics are designed with the principal 'How are they changing the system in the outside world?'

See David Kryl's slides here.

This methodology has not yet been trialled in research for international development.

RAPID Outcome Assessment

RAPID Outcome Assessment (ROA) is a learning methodology to assess and map the contribution of a project's actions to a particular change in policy or the policy environment. It is a flexible tool which can be used alongside other evaluation tools and methods. The framework was developed by the CGIAR-affiliated International Livestock Research Institute and the Overseas Development Institute's (ODI) Research and Policy in Development (RAPID) programme, in response to the need to better understand the processes and mechanisms that lead to pro-poor decisions at the policy level. It combines elements from the Outcome Mapping approach (examines the progressive behaviour changes amongst the key actors that the programme is directly influencing), Episode Studies (tracks back from a policy change to identify the factors that contributed to it), case study analysis (tracking forwards from research), and Most Significant Change (helps to identify and prioritise the key changes). ROA is designed to assess the contribution of research both during and after a project is completed.

See John Young's slides.

Contribution analysis

Contribution analysis can be used as a way of linking activities to outcomes by creating convincing evidence chains that would demonstrate research uptake, use and impact. It offers a step-by-step approach designed to help managers, researchers, and policymakers arrive at conclusions about the contribution their programme has made (or is currently making) to particular outcomes.

A logic-model is developed, and the assumptions underlying the model are made explicit. Evidence is then gathered to test the validity of these assumptions.

Contribution analysis allows for the analysis of process and outcomes and

acknowledges that there are many factors influencing change. It provides evidence to demonstrate the outcomes from research and knowledge exchange and can help with planning knowledge exchange activities and stakeholder engagement. It is adaptable to different timeframes, and can assess immediate or intermediate impact. It can be used for planning and evaluating in real time or retrospectively.

Contribution analysis is still a new method however, with limited examples of usage. The logic-model approach looks and feels linear (although it does allow for non-linearity). The focus on positive outcomes, or a desired impact might make it difficult to use where the outcome of research cannot be predicted.

Adapted from Morton, S. (2012). Exploring and Assessing Research Impact. Social Policy. Edinburgh, University of Edinburgh. PhD.

See here for more information, or these resources (1, 2, 3, 4).

Outcome Mapping

This approach is a participatory planning tool which unpacks an initiative's Theory of Change, provides a framework to collect data on immediate, basic changes that lead to longer, more transformative change, and allows for the plausible assessment of the initiative's contribution to results via 'boundary partners'. It has a number of strengths

- Assesses both process and outcome
- Assesses contribution (not attribution)
- Enables stakeholder dialogue, learning and consensus building, as well as accountability
- Captures elements of policy implementation rather than focusing on 'paper' policy

And weaknesses:

- Does not evaluate value for money
- Observed behavioural change posits a link between cause and effect which may be impossible to demonstrate
- Costly and time-intensive

Some experts at the workshop considered that Outcome Mapping tends to undervalue the role of research in causing change.

Process tracing

Process tracing is the systematic examination of diagnostic evidence - often understood as part of a temporal sequence of events or phenomena - selected and analysed in light of research questions and hypotheses posed by the investigator. Process tracing can contribute decisively both to describing political and social phenomena and to evaluating causal claims.

Process tracing can be summarised in terms of four empirical tests,

- 'straw in the wind', which lends support for an explanation without definitively ruling it in or out,
- [jump through the] 'hoop', which rules out an explanation if it fails this test,
- 'smoking gun', which rules in an explanation if it is present, and
- 'doubly definitive' where the 'clue' is both necessary and sufficient support for the explanation.

The tests are classified based on two criteria: whether passing the test is necessary for establishing a causal connection, and whether it is sufficient for establishing a causal connection. Process tracing can be used both to see if results are consistent with the Theory of Change and to see if alternative explanations can be ruled out.

More information is available here.

Oxfam GB noted three lessons learned from using process tracing as part of their evaluations for projects under their 'Citizen Voice' and 'Policy Influencing' thematic areas: drawing out the (often implicit) Theory of Change was slow and time consuming; identifying 'the right' interim outcomes to focus on was difficult; and, specific to process tracing

"In the absence of a "signature" – something that unequivocally supports one hypothesized cause – what constitutes credible evidence of the intervention's contribution to policy change? Can we overcome the charge of (positive) bias so often levelled at qualitative research?"

Oxfam's draft protocol on process tracing for evaluating 'Citizen Voice' and 'Policy Influencing' interventions.

Most Significant Change

This participatory approach collects and analyses personal accounts of change, and includes processes for learning about what changes are most valued by individuals and groups. Significant change stories emanating from the field level are collected,

and the most significant of these stories are systematically selected by panels of designated stakeholders or staff.

See these webpages (1, and 2) for further resources on the methodology, and this ODI toolkit briefing. Most Significant Change has been successfully used in many different types of programmes, and in many different countries around the world, since its development in the mid 1990's. Evaluation of development-relevant research has included the CGIAR Challenge Program on Water and Food.

Like all research methods, it has certain limitations and issues that need to be taken into account. Biases that have been associated with Most Significant Change include:

- Biases towards stories of success – this can be overcome by creating a specific domain to capture negative stories.
- Biases towards the views of those who are good at telling stories - a good reason for not using Most Significant Change as a stand-alone tool for M&E.
- Subjectivity in the story selection process. This process tends to reflect the values of the people on the selection panels. However, recording the reasons for selecting stories helps to strengthen the process.

Most Significant Change also raises the important issue of voice and power associated with who participates in the story selection process. This can be overcome by ensuring that the people on the selection panels represent a wide range of backgrounds and values. There are also potential issues with anonymity and confidentiality of data and ethical concerns with obtaining informed consent for the use and distribution of the stories collected.

Social network analysis

Social network analysis aims to understand the actors and the relationships between them in a specific social context. These focuses help to understand how the actor's position in a network influences their access to resources such as goods, capital and information.

Social network analysis is often applied to identify both information flows and bottlenecks. It uses a variety of tools – network diagrams, network matrices and mathematical measures – to depict and aid understanding of social networks. The key distinctive feature of social network analysis is a focus on relationships between actors, rather than their individual characteristics.

New Directions for Evaluation commissioned a special issue on Social network analysis in 2005. To date, it has been used for research priority setting, strengthening partnerships, communication and fund-raising for research institutions (see here, or

here for a wide range of modelled examples and live examples) but not to evaluate the impact of research.

Actor Network Theory

Actor Network Theory (ANT) is a conceptual frame for exploring collective sociotechnical processes. ANT attempts to "open the black box" of science and technology by tracing the complex relationships that exist between governments, technologies, knowledge, texts, money and people. ANT challenges some common epistemological convictions by rejecting essential subject/object, culture/nature or society/technology distinctions. Entities, whether people or technologies, are not fixed and do not have significance in and of themselves. Instead, they achieve significance through relations with other entities.

It can be used as a way of understanding how research has impact, and it emerged over the past twenty years as a major conceptual force in social science. To date it has not been used much in international development, and is yet to be used in evaluating the impact of research. See this website for presentations from a workshop on ANT in international development.

Qualitative Document Analysis

Qualitative Document Analysis (QDA) is the range of processes and procedures used to move from the qualitative data that have been collected into some form of explanation, understanding or interpretation of the people and situations under investigation. QDA is usually based on an interpretative philosophy: the idea is to examine the meaningful and symbolic content of qualitative data.

The methods can be applied to a wide range of sources (including policy statements, technical reports, minutes, speeches), and are good at providing contextual understanding. In addition, they are potentially cost-effective. However, QDA relies upon the quality of existing records and access to these.

These tools (like literature review) are used routinely in the evaluation of research programme impact in development.

Surveys

A pre-formatted series of questions asked of multiple actors, generating both quantitative and qualitative data.

The strengths and weaknesses of surveys for evaluating impact are relatively well

known.

Strengths:

- Cost-effective means of obtaining data from a range of actors
- Allows wider involvement of stakeholders than possible with interviewing
- Can identify greater breadth of outcomes than other methods
- Useful for the triangulation of findings from other methods

Weaknesses:

- Relies upon access to populations of interest, or robust samples thereof
- Reflects biases in those selected and those who respond (can suffer from low response rates)
- Relatively unresponsive to unforeseen issues
- May require follow-up interviews to fully understand the results
- A questionnaire approach can overlook nuances and subtleties surrounding influence

Survey based evaluation methods have been used extensively to assess the impact of international development interventions, including research programmes. Surveys are predominantly used as part of mixed-methods evaluations.

Cohort studies

The Wellcome Trust profiled their Career Tracker – a prospective system to track the career choices and progression of cohorts of individuals funded by the Trust, over time. The Trust believes that developing individuals, and building research capacity is vital: this tracker is one tool to try to assess whether the funding mechanisms work for the future of science.

The Trust now have three years' worth of data and are already starting to see trends and differences across cohorts. They have managed to secure very high response rates. However, the resources needed to set-up the trackers and keep responses high are relatively intense. In addition, the questionnaire has to be relatively high level to keep it short and allow benchmarking between cohorts – this limits the detail the method can extract.

Another concern is that the cohorts may be responding out of goodwill to their funder – when they are no longer Wellcome funded, response rates may drop decreasing the

power of the approach.

See Bryony Rayfield's slides here.

Randomised Controlled Trials

RCTs are an approach that produces an estimate of the mean net impact of an intervention by comparing results between a randomly assigned control group and experimental group or groups.

There is a very vigorous debate about the role of RCTs in impact assessment of development interventions. See here, here, and here among numerous other opinions and resources. However, there have been no attempts yet to apply RCTs to evaluate the impact of research at the programme level in development.

Alternative indicators

A selection of new alternative indicators have been developed recently in response to concerns of slowness, narrowness of breadth, or improper implementation with traditional indicators like journal impact factors. 'Altmetrics' purport to widen the reach of academic reward systems to account for the diverse ways in which scholars actually communicate – blog discussions, Twitter citations, and open dataset repositories are just a few examples. Altmetrics could therefore usefully measure factors related to conceptual impact - see Why and what? Motivations for evaluating impact and tools for framing the right question. This fast moving field includes aggregators of digital coverage (e.g. Twitter, Facebook and blogs) or domain specific impact measuring software packages. See these two helpful reviews (1, 2).

Altmetrics may have more value in highlighting the varying routes to conceptual and instrumental impact, and monitoring activity, rather than in the specific quantification of impact they provide. A lot of the indicators measured by altmetrics have very short half-lives – and so could potentially be registering hype/buzz as opposed to more subtle longer term change.

There are examples of trying altmetrics in research for international development: ODI bring a number together into a dashboard to monitor and evaluate the impact of their work.

For an example from the Wellcome Trust, see Kevin Dolby's slide here.

Digital footprint analysis

A group at the LSE created a dataset of 350 randomly chosen UK social science academics. They looked at what they published, their academic citations and their external references – creating a digital footprint of their work using large scale, unobtrusive measures.

They studied the way the work was cited and referenced by civil society, government and business, both in the UK and globally.

See Jane Tinkler's slide here and further information on the technique here. This sort of analysis has not yet been applied in research for international development.

SenseMaker

SenseMaker is a software tool that helps to find patterns among stories (fragments of information about a complex change process rather than whole case studies) – for use as an evaluation tool. After collecting these fragments of information in a raw form, the person who provided them decides what they mean (as opposed to computer based analysis or expert judgement) by assigning them to a location on a 3x3 spectrum. This gives quantitative data – which can be visualised and analysed.

See this video clip for an explanation of the tool. It has been used a number of times in evaluating development interventions.

Recipient led impact

IDRC are trialling new approaches to evaluation that are 'recipient led' to try and ensure the evaluation reflects the issues, needs and perspectives of that country/domain. So, for example, an evaluation led by a Ministry of Health, Education, or Agriculture, would seek to assess to what extent the research has generated knowledge to allow that Ministry to design and implement policy. This aims to shift the evaluation focus from the programme and how it fits the goals of the funding agent, to the development priorities, policies and practices in the country in question, and to evaluate the research outputs against the priority questions they think need to

be addressed.

This approach is an evolution of participatory evaluation: a range of approaches that engage stakeholders (especially intended beneficiaries) in conducting the evaluation and/or making decisions about the evaluation.

What to choose?

The range methods and approaches profiled on this page have different strengths and weaknesses. Some have been used to evaluate the impact of research, some have not. Some have been tested in international development, some have not.

The authors of these resources are not aware of a framework for choosing between methods. To begin thinking about a framework, and to inform others in the field, we are keen to collect and profile examples of how evaluators have chosen between methods.