Contribution Analysis and Bayesian Confidence Updating

A brief introduction

This paper provides an introduction to contribution analysis and outlines how it establishes whether or not an intervention or investment contributed to change. It also introduces Bayesian Confidence Updating, and explores how this might be combined with contribution analysis to provide a robust assessment of CDC's impact on mobilisation without the use of a counterfactual.

What is contribution analysis? Contribution analysis is a theory-based evaluation approach that provides a systematic way to arrive at credible causal claims about an intervention's contribution to change.[1] In a nutshell, it involves developing and assessing the evidence for a logic model or theory of change (ToC), in order to explore the intervention's contribution to observed outcomes. By verifying the ToC that the intervention is based on, and taking into consideration other factors that may have influenced outcomes, contribution analysis can provide evidence that the intervention did or did not make a difference.[2] Box 1 presents an overview of the steps involved.

Contribution analysis allows a robust assessment of cause and effect when it is not practical to design an experiment to measure the attribution of a particular change to a particular intervention. It also recognises that an intervention might only be one of a number of factors contributing to the observed effects and provides a framework to assess the extent that an intervention has contributed while recognising other influencing factors.

> *"Contribution analysis recognises that an intervention might only be one of a number of factors contributing to the observed effects"*

Origins of contribution analysis. Contribution analysis was developed by John Mayne, as an approach to assess Canadian public sector interventions.[3] Over the years his ideas have influenced many other fields, and the approach has attracted widespread attention within the global evaluation community.[4]

How does contribution analysis establish a causal link? Contribution analysis seeks to increase our confidence that an intervention contributed to an outcome. It demonstrates that contribution is *probable* through

Box 1. The six steps of contribution analysis

Step 1: Set out the-cause effect issue to be addressed. Scope the specific cause-effect question that is being asked, and determine the expectations of commissioners and the level of confidence needed in answering the question. Conduct initial thinking about other key influencing factors, and how plausible the expected contribution is given the nature and size of the intervention.

Step 2: Develop the ToC. Set out the postulated ToC for the intervention, including the underlying assumptions, risks, unintended effects and other explanatory factors.

Step 3. Gather existing evidence on the ToC. Including existing intervention data and evidence from previous evaluations and relevant research.

Step 4: Assemble and assess the contribution story, and challenges to it. Use existing evidence to 'assemble the contribution story' – evidence on the results, assumptions and influence of other factors. Use this to assess strengths and weaknesses in the ToC, and the relevance of other factors.

Step 5: Seek out additional evidence. Determine what additional evidence is needed to strengthen the contribution story, and gather new evidence.

Step 6: Revise and strengthen the contribution story. Use new evidence to revise the contribution story and reassess its strengths and weaknesses, along with the relevance of other factors.

Source: Mayne 2013[3]

building a well-evidenced case that the intervention contributed, iteratively collecting and analysing additional evidence over time to revise and strengthen the contribution story.

There are four conditions needed to infer causality in contribution analysis.[5]

1. **Establishing that the ToC is plausible.** The results chain and assumptions are plausible, sound, informed by existing research, and supported by key stakeholders.
2. **Establishing intervention fidelity to the ToC**. The intervention was implemented as planned.
3. **Verifying the ToC**. The chain of expected results occurred, and the causal assumptions held.
4. **Accounting for other influencing factors.** Other potential causes of the results have been assessed and their relative contribution recognised.

These conditions are premised on a *generative* logic of causality – the idea that causal links can be demonstrated through a fine-grained explanation of how and why the intervention caused the outcome.[6] We use this form of causality frequently in everyday life. For example, when we turn on a light switch we expect the light to come on (a cause-effect relationship). If this doesn't happen, we diagnose the problem through investigating the different parts of the 'causal chain' in order to assess how and why the switch failed to work – opening up the switch to check the wiring, changing the bulb etc.[7]

"Generative logic of causality involves the idea that causal links can be demonstrated through a fine-grained explanation of how and why the intervention caused the outcome."

Contribution analysis therefore explains intervention impact in a different way to experimental and quasi-experimental evaluation approaches (which are underpinned by correlational or counterfactual models of causality).[8] Counterfactual approaches assess intervention impact by comparing observed outcomes to what would have happened in the absence of the programme. Mechanism-based (generative) approaches assess impact by investigating the causal processes that generate the impact.[9] See Box 2 for a further example.

Is contribution analysis robust? Contribution analysis provides a helpful and intuitive framework, but not a prescribed set of methods and processes for assessing a ToC. In practice, examples of contribution analysis can fall down somewhat when it comes to robustly establishing the strength of the causal links in the ToC and the importance of other factors. For example, Schmitt & Beach criticise the approach for 'relegating' causal processes to assumptions, which are frequently not studied empirically and therefore are not tested in a robust way.[10] Befani and Stedman-Bryce argue that contribution analysis, alongside other theory based approaches, fails to provide sufficient guidance on how to collect data or how to assess that data in order to support a ToC. This means it isn't always possible *"to link the observations made during the data-collection process with the contribution claim; or to understand why the contribution was considered 'significant' as opposed to 'fair' or 'small' or 'non-existent.'"*[11]

Box 2. Counterfactual vs generative approaches to causality: an example

Consider the example of a newly developed drug to treat anxiety. The effectiveness of the drug is likely to be established through a randomised controlled trial, which compares outcomes for randomly assigned people who received the drug with outcomes for those who didn't. This will establish the extent to which the drug is effective using a counterfactual model of causality. However, the trial will have been preceded by many years of basic research, using generative causality to develop and test theory about how and why a specific formulation is expected to make a difference – understanding the biological, chemical and psychological mechanisms that will lead to the drug improving anxiety.

Source: Pawson (2017)[10]

The promise of Bayesian Confidence Updating. In recent years, evaluators have been working on innovative approaches to robustly establishing the strength of causal links in theory-based evaluation. This includes exploring the potential of process tracing to strengthen contribution analysis. Like contribution analysis, process tracing is a theory-based evaluation approach, but it involves a much more specific and transparent approach to assessing the strength of evidence behind causal claims. One technique connected to process tracing has proved very useful: Bayesian Confidence Updating.[13] This involves assigning probabilities to causal claims or hypotheses, and formally weighing evidence to update the researcher's confidence in these claims being true or false. This technique draws upon Bayesian statistics, which uses Bayes' theorem[14] to update the probability for a hypothesis as more evidence or information becomes available. A worked example is provided in Box 3 below.

Box 3. Worked example of developing and assessing a contribution claim

This box provides a simplified worked example, based on a case study of CDC's investment in the Global Environment Facility's Africa Sustainable Forestry Fund (ASFF) based on CDC's Annual Review (CDC 2010). In 2008, CDC issued a request for proposals for a forestry-focused fund in Sub-Saharan Africa, and selected a fund manager. In 2010, CDC provided cornerstone investment to get the fund started. The fund was able to attract substantial further investment and is shared as a *'good example of how CDC's robust approach and strong reputation can catalyse other investors to commit their capital'.* How might Process Tracing principles be applied in this case, to robustly assess CDC's contribution?

Step 1: Develop a contribution claim

An initial contribution claim is developed, based on a review of available documents.

CC1.CDC's anchor investment in ASFF attracted other investors owing to CDC's robust approach and strong reputation.

Through consultation, the claim is then further refined, to provide more detail about precisely how CDC's investment is thought to have attracted other investors. This will incorporate the behavioural 'COM-B' framework, reflecting on how CDC has influenced capability, opportunity and/or motivation in order to influence behaviour. Also, because we are interested in both *how* CDC contributed and the *strength of this contribution*, the statement is divided into two claims to be tested, such as:

CC2.Part I. CDC's anchor investment in ASFF <u>reduced</u> the perceived risk of investing in the forestry sector among investors (DEG, IFC, Proparco, FinnFund, MAEC), owing to CDC's robust approach and strong reputation. As a result, these investors were motivated to commit capital to the fund.[1]

CC3.Part II. CDC's anchor investment in ASFF was <u>the most important</u> driver of reducing perceived risk. It was more important than activity Y, and activity Y was more important than activity Z (but all had some influence).

The evaluators then determine the prior probability of CC2 and CC3 being true, before any evidence is collected. With no prior information about the claim, and no reason to believe it is more or less valid, the evaluators might set the prior probability at 0.5, which is the 'no information' situation in Bayesian statistics.

Step 2: Design data collection

The evaluators consider what data they need to collect to evaluate the contribution claim, thinking about the probative value of particular types of evidence and prioritising evidence with higher probative value. Particular attention is paid to 'expect to see' evidence and 'smoking gun' evidence.

<u>Examples of 'expect to see' or 'hoop' evidence:</u> This is evidence that must be found in order to keep *CC2* under consideration. For example:

E1 [Monitoring/media?] data confirms that DEG, IFC, Proparco, FinnFund and MAEC made limited (or considerably smaller) investments in the forestry sector prior to their investment in ASFF.

This is 'expect to see' evidence because: if the investors were already investing significant amounts in the forestry sector before ASFF was established, it does not make sense to claim that ASFF reduced the perceived risk of investment.

[1] Note, this example is relatively simplistic for demonstration purposes – in practice, the contribution claim may be unpacked into several linked hypotheses.

Box 3 Continued

'Smoking gun' evidence: This type of evidence is harder to find but is the 'holy grail' for evaluators. It has the ability to dramatically increase the evaluator's confidence in a hypothesis, because it is very unlikely that the evidence will be found unless the hypothesis is true.

E2. Written evidence (e.g. meeting minutes) from DEG, IFC, Proparco, FinnFund and/or MAEC investment committees link decisions to invest in ASFF to CDC's anchor investment reducing the risk of investing in forestry.

This is smoking gun evidence because: it is unlikely that written documentation would cite the influence of CDC and link it to the reduction in risk, unless this was in fact true.

'Straw in the wind' evidence: This type of evidence is not enough on its own to prove the causal claim but can increase the evaluator's confidence in the claim when considered alongside other independent sources of evidence.

E3. Investors in ASFF confirm that historically they have perceived the forestry sector as a risky area for investment, until CDC invested.

E4. Investors in ASFF confirm that CDC is perceived as having a robust approach and strong reputation and that this influenced their decision to invest.

Each of these pieces of evidence, on their own, are not enough to confirm the contribution claim – as investors may have various incentives to exaggerate CDC's influence, or may overstate it because of confirmation bias. However, if both are found together, or in combination with E1 and/or E2, this helps increase confidence in the contribution claim.

Step 3: Conduct data collection and weight evidence

Data collection is conducted, and analysis confirms whether evidence E1–E4 was found. During analysis, the evaluators then determine the sensitivity and type 1 error values for specific pieces of evidence.

5. **Sensitivity:** The probability of finding evidence x if CC2 is true.
6. **Type 1 error:** The probability of finding evidence x if CC2 is false.

Each of these values can be quantified, as a subjective probability between 0 and 1. For example, E1 might be assigned a sensitivity value of 0.95, as it is very likely that investment in forestry prior to ASFF was considerably smaller, if CC2 is true. The evaluators might assign a type 1 error of 0.6, because it is also more likely than not that investments prior to ASFF were smaller even if CC2 is false – as investors may have been attracted to invest in ASFF for some reason other than CDC reducing the perceived risk.

After assigning these values, the evaluators then apply Bayes' formula to each piece of evidence, in order to calculate the posterior probability of CC2 being true. From a starting point of 0.5 ('no information about whether CC2 is true'), applying Bayes' formula will provide a new probability based on the evidence collected. If we find evidence E1, and apply Bayes' formula (with a prior of 0.5, a sensitivity of 0.95 and a type 1 error of 0.6), this results in a posterior of 0.61 – in other words, this has increased our confidence (but not hugely) that CC2 is true. If other pieces of evidence (E2–E4) are also found, the calculations can be combined to further increase our confidence in the contribution claim.

Step 4: Put the claim and findings up for challenge.

The value of the contribution tracing process rests on the validity of the subjective probability estimates, particularly around sensitivity and type 1 errors for given pieces of evidence. Step 4 therefore involves a constructive conversation with key stakeholders, including other members of the evaluation team and CDC stakeholders knowledgeable about the case under investigation, in order to discuss, challenge and reach consensus on the probative value of evidence. This step might be conducted after the preceding three, but where possible should take place earlier to help identify 'expect to see', smoking gun and straw in the wind evidence in advance of data collection.

- Bayesian Confidence Updating emphasises the *probative value* of evidence, not the *amount* of evidence. The probability of a causal hypothesis being true is not proportional to the amount of evidence collected. The evaluators might collect a lot of evidence that has little probative value (for example many interviews with intervention staff or direct beneficiaries claiming the influence of the intervention) – this may be less informative than just one piece of evidence with high probative value (such as a civil servant with no stake in the intervention stating the impact of a piece of funded research). Thinking about probative value encourages evaluators to identify and prioritise collecting the most useful and insightful pieces of evidence, rather than a specific quantity of evidence.

> *"The probability of a causal hypothesis being true is not proportional to the amount of evidence collected"*

- It forces the evaluators to make their assumptions about evidence clear and transparent, so that others can interrogate and dispute them. For example, we might assume that a civil servant is unlikely to verify the influence of a research report on a policy unless this did in fact happen (E^4 in Box 3). However, especially in an international development context, there might be various incentives for civil servants to exaggerate the impact of the report (for example the desire for donor funding). The process of assigning probabilities helps surface these kinds of issues.
- It encourages systematic consideration of validity and reliability – ensuring that weaknesses in the evidence are directly reflected in the final results (as they will be reflected in the probabilities assigned), rather than hiding in the 'limitations' section of a report, never to be mentioned again.

In summary, contribution analysis holds significant promise as an approach that can robustly assess an intervention's impact when it is not practical to design an experiment, drawing on a generative model of causality. One of its major advantages is in recognising that an intervention might only be one of a number of factors contributing to the observed effects, and providing an intuitive framework to assess the extent that an intervention has contributed while recognising other influencing factors. However, one weakness in the approach is that it does not provide much detail on how to collect and assess evidence to robustly establish the strength of the causal links in the ToC. Bayesian Confidence Updating can help address this challenge, providing a specific and transparent approach to assessing the strength of evidence underpinning a ToC.

Notes:

[1] Mayne, John. 2012a. "Contribution Analysis: Coming of Age?" *Evaluation* 18 (3): 270–80. doi:10.1177/1356389012451663.

[2] Mayne, John. 2008. "Contribution Analysis: An Approach to Exploring Cause and Effect." *International Learning and Change (ILAC) Brief*, ILAC Brief, 16.

[3] Key papers include: Mayne J. Addressing attribution through contribution analysis: Using performance measures sensibly. Can J Progr Eval. 2001;16(1):1-24. doi:10.1007/978-94-007-6386-9_47. Mayne J. Contribution analysis: An approach to exploring cause and effect. *Int Learn Chang Br*. 2008;16. Mayne J. Addressing Cause and Effect in Simple and Complex Settings through Contribution Analysis. In: Schwartz R, Forss K, Marra M, eds. *Evaluating the Complex*. Transaction Publishers; 2011.Mayne J. Contribution analysis: Coming of age? *Evaluation*. 2012;18(3):270-280. doi:10.1177/1356389012451663.

[4] See e.g. Kotvojs F, Shrimpton B. Contribution analysis: A new approach to evaluation in international development. Eval J Australas. 2007;7(1):27-35. Dybdal L, Nielsen SB, Lemire S. Contribution analysis applied: Reflections on scope and methodology. Can J Progr Eval. 2010;25(2):29-57.

[5] Befani, Barbara, and John Mayne. 2014. "Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation." IDS Bulletin 45 (6): 17–36. doi:10.1111/1759-5436.12110.

[6] Mayne, John. 2012b. "Making Causal Claims." International Learning and Change (ILAC) Brief 26.

[7] Adapted from Rothman, K. (2005) Causation and Causal Inference in Epidemiology, pp 144-150 of the *American Journal of Public Health*, Supplement 1, Volume 95.

[8] One of the most well-known applications of generative causation is realist evaluation, which studies the causal 'mechanisms' (or change processes) that explain how and why interventions lead to change (Pawson, Ray, and Nick Tilley. 1997. *Realistic Evaluation*. London: Sage.)

[9] Shaffer, Paul. 2014. "Two Concepts of Causation: Implications for Poverty." Development and Change 46 (1): 148–66. doi:10.1111/dech.12140.

[10] Schmitt, Johannes, and Derek Beach. 2015. "The Contribution of Process Tracing to Theory-Based Evaluations of Complex Aid Instruments." doi:10.1177/1356389015607739.

[11] Befani, B., and G. Stedman-Bryce. 2016. "Process Tracing and Bayesian Updating for Impact Evaluation." Evaluation 1 (19). SAGE Publications. doi:10.1177/1356389016654584.

[12] This example was presented by Ray Pawson (one of the founders of realist evaluation) in an Exaugral Lecture at Leeds. Pawson argued that pharmaceutical RCTs are only possible and meaningful because of many prior years of basic science, which involves theory building and testing through generative causation. https://realism.leeds.ac.uk/ray-pawson-exaugural-lecture/

[13] Befani, B., and G. Stedman-Bryce. 2016. "Process Tracing and Bayesian Updating for Impact Evaluation." Evaluation 1 (19). SAGE Publications. doi:10.1177/1356389016654584.; D'Errico, Stefano, Barbara Befani, Francesca Booker, and Alessandra Guiliani. 2017. "Influencing Policy Change in Uganda- Impact Evaluation." http://pubs.iied.org/G04157/.

[14] Bayes' theorem, based on the work of Thomas Bayes (1701–1761).

Brief written by Mel Punton and Chris Barnett, Itad Ltd.

We want the resources invested in international development to deliver the best possible results for the poor.

Through our innovative consultancy services in monitoring and evaluation we provide the insight and ideas to ensure that they do.

Itad

Preece House
Davigdor Road
Hove BN3 1RE
United Kingdom

T +44 (0) 1273 765 250
E mail@itad.com
W itad.com