Causal mapping in evaluation

CONTENTS

Intro

> Causal mapping has been used in many different fields. In this chapter we look at how it can be applied in evaluation; its strengths and weaknesses.

Causal mapping is particularly useful for evaluations that focus on learning to inform program improvement: visual representation of causal links between context, activities and outcomes can help to facilitate the sharing and collaborative use of findings.

Causal mapping can be used during a program lifespan to inform adaptive management and as part of a final evaluation.

> From (Powell et al., 2024)

Causal mapping offers ways to organise, combine, present and make deductions from a large number of relatively unstructured causal claims – the sort of data that are often collected in evaluations.

References

> Powell, Copestake, & Remnant (2024). *Causal Mapping for Evaluators*. https://doi.org/10.1177/13563890231196601.

Causal mapping can complement contribution analysis

Contribution analysis is about testing and refining a theory of change to build a credible case for contribution. It's focused on a specific theory of change and on the contributions to outcomes. These contributions might happen along a causal chain, but CA tends to be less explicit about how they are to be traced.

Causal mapping is strongly related. It is not an evaluation method in its own right, but more of a tool which can assist with CA. it's about visualizing and interrogating the whole web of causes and effects. It explicitly addresses the challenge of overlap and influence between causes and effects. It doesn't rely on a fixed theory of change but on the factors and links between them which are actually mentioned in documents and interviews.

How causal mapping can help with Contribution Analysis:

- by helping to assemble all the relevant evidence along some causal pathway or pathways from intervention to outcome. Causal mapping has a strong understanding of chains and transitivity. Contribution Analysis can then focus on what it is best at, namely weighing up different explanations for an outcome and how much our intervention really contributed to it.

Causal mapping can complement Outcome Harvesting

Outcome harvesting is about collecting and explaining a (hopefully long and substantial) list of intended and unintended outcomes after the fact, and identifying how the programme contributed to the outcomes. But it's just like a list of cause-effect relationships. It can be a challenge to understand how those causes and effects overlap with one another or influence one another.

Causal mapping is quite similar. It is not an evaluation method in its own right, but more of a tool which can assist with either. it's about visualizing and interrogating the whole web of causes and effects. It explicitly addresses the challenge of overlap and influence between causes and effects. It doesn't rely on a fixed theory of change but on the factors and links between them which are actually mentioned in documents and interviews.

How causal mapping can help with Outcome Harvesting:

- causal mapping can input a pile of Outcome Harvesting data and link them all together into the form of a larger web. (Britt et al., 2025).

The result of an evaluation is a qualitative causal model

> From (Powell et al., 2024)

Causal mapping has some limitations. First, the credibility of the causal arguments which can be derived from a map is limited by the credibility of the original data sources. We see the job of causal mapping as collecting, organising and synthesising a large number of claims about what causes what; drawing conclusions about what this actually reveals about the world is a final step that goes beyond causal mapping per se. In specific cases, establishing explicit and context-specific rules of inference may help to make this final step.

[!01381 Rubrics can help make evaluative judgements from causal mapping](#)

References

Powell, Copestake, & Remnant (2024). *Causal Mapping for Evaluators*. https://doi.org/10.1177/13563890231196601.

The elephant in the room -- causal inference

[Responding](#) to our [one-page description of causal mapping](#), Julian King says the elephant in the room with causal mapping is: can causal mapping really help you get from causal opinions to causal inference?

The short answer is: sure it can help you, the evaluator, make that leap. But it, causal mapping, does not give out free passes.

But in more detail, here are four more responses to the elephant.

1: **Causal mapping ignores the elephant. On its own, causal mapping doesn't even try to warrant that kind of step**: it humbly assembles and organises heaps of assorted evidence in order for the actual evaluator to make the final evaluative judgement. Unlike evidence from an interview, or a conclusion from process tracing or from a randomised experiment, causal mapping evidence isn't a *kind* of evidence, it's an *assemblage* of those other kinds of evidence. It certainly isn't a shortcut to get cheap answers to high-stakes answers by conducting a few interviews with bystanders.

If you have to answer high stakes causal questions like "did X cause Y" and "how much did X contribute to Y" and have just a handful of pieces of evidence, there isn't much point using causal mapping. Causal mapping is most useful for larger *heaps* of evidence, especially from mixed sources and of mixed quality; it gives you a whole range of ways of sorting and summarising that information, on which you can base your evaluative judgements. What it doesn't give you is a free pass to any evaluation conclusions, and especially not the high stakes ones which occupy so much of our attention when we think and write about evaluation.

2: In most actual causal mapping studies, **the elephant usually doesn't even enter the room**. Usually, we aren't dealing with monolithic, high-stakes questions. Most causal mappers are looking for (and finding) answers to questions like these:

- In which districts was our intervention mentioned most often?
- Do children see things differently?
- How much evidence is there linking our intervention to this outcome?
- Does our project plan see the world in the same way as our stakeholders?

All of these are relevant questions for evaluations. Some of them might feed into judgements about relevance, or about effectiveness or impact, and so on. We might notice for example that there is some evidence for a direct link from an intervention to an outcome, and much more indirect evidence, and some of those paths remain even when we remove less reliable sources. We can even compare the quantity of evidence for one causal pathway with the quantity of evidence for a different pathway. We can ask how many sources mention the entirety of a particular pathway, or we can ask which pathways have to be constructed out of evidence from *different* sources. (On the other hand we don't, for example, make the mistake of inferring from the fact that there is a *lot* of evidence for a particular causal link that the link is a *strong* one.)

All of this is bread and butter for evaluators, even though it doesn't answer those elephant questions.
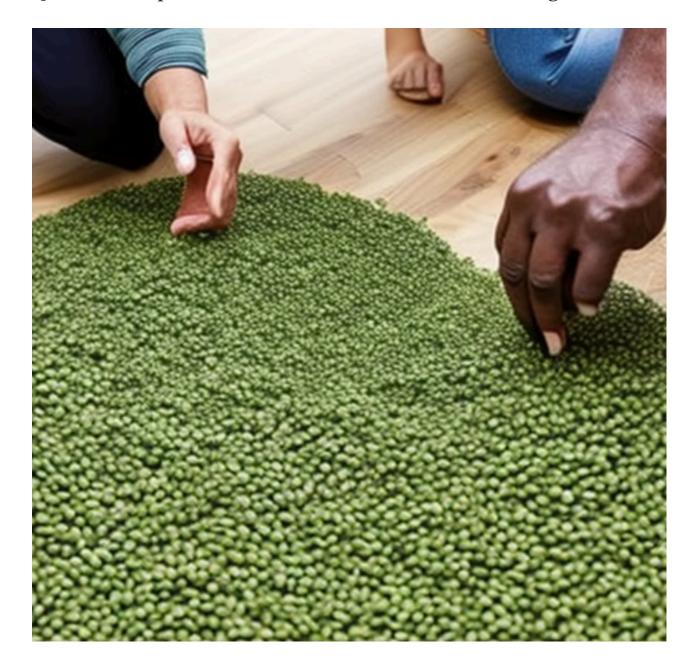
3: Causal mapping pushes back against the elephant.

In *every* evaluation, the evaluator assembles some evidence and makes an evaluative judgement on the basis of it. **All evaluation involves causal mapping** in this sense. Occasionally there is, or seems to be, only a single piece of evidence in the heap – perhaps, evidence from a controlled experiment. But the final judgement is the evaluator's responsibility, and (perhaps implicitly) must take into account other factors: "this is a controlled experiment, it was carried out by a reputable team, ... but wait, their most recent study was criticised for allowing too much contamination .... but wait, the effect sizes were calculated with the latest method and controlled experiments seem to be a good way of reaching causal conclusions ...", and so on. An essential part of the evaluative process is also careful consideration of how exactly to formulate a conclusion, bearing in mind the context and the audience and how it will be generalised and applied. So, in practice, there is always a heap of factors to consider, often involving different parts of more than one causal pathway, even when the heap seems to be dominated by one or two elephants.

4: Causal mapping embraces the elephant. In most causal mapping studies, we do not in fact simply assemble the evidence we already have but actively gather it systematically. A good example is QuIP, the Qualitative Impact Assessment Protocol. The evidence is "only" the considered opinions of carefully selected individual stakeholders, but it is gathered using blindfolding techniques to minimise bias so that, once assembled and organised with causal mapping, the evaluative leap from opinions about causality to conclusions about causality can be

made with more confidence, transparently and with appropriate caveats. Still, it's not the causal mapping itself which makes or warrants the leap, it's the evaluator, using evaluative judgement.

Qualitative impact evaluation is less interested in the strength of effects

Weird image of people counting beans generated by Canva's AI

Soft versus hard impact evaluation approaches? Quant versus Qual? Is there an essential difference?

Summary: quant and qual impact evaluation approaches are different ballparks because quant approaches attempt to estimate the *strength* of causal effects. Whereas qual approaches either don't use numbers at all or only do calculations about the *evidence for the effects*, not about the effects themselves and in particular we don't estimate strength of effects.

Here's the question: how can we distinguish "soft" approaches to impact evaluation like Outcome Harvesting, QCA, causal mapping, Process Tracing, Most Significant Change, Realist Evaluation and so on from statistics-based causal inference (SEM, DAGs, RCTs etc)?

Here are two bad answers:

- We can't distinguish our "soft" approach(es) by saying that we attempt to assess causal contribution and answer questions about for whom and in what contexts etc, because quantitative approaches attempt all of that too.
- We can say that we are focused on complex contexts, but there's nothing to stop someone using say OH in a non-complex context either is there? In any case whether a context is complex or not is also a matter of how you frame it, no? And there's in fact no shortage of examples where quant approaches have been used in complex contexts.

Here's a better answer: these "soft" methods are qualitative, in the sense that where we involve numbers at all, our arithmetic is essentially an arithmetic of *evidence for* causal effects: is there any evidence for one pathway, how much, how much compared with another? For example, Process Tracing sometimes does calculations about the relative certainty of different causal hypotheses. QCA counts up configurations.

Whereas quant causal analysis involves estimating the *strength* of causal effects (as well as having clever ways to reduce the bias of those estimates).

As far as I know, qualitative approaches never attempt this (calculating the strength of a causal effect). We might conclude that the evidence suggests a particular effect is *strong*, for example because we have collected and verified *evidence for a strong connection*. But we don't, say, combine this with another set of evidence for a very weak connection and conclude that the strength of the effect was only moderate (we don't do maths on the strengths).

It's true that qual approaches also do causal inference in the sense of making the jump from evidence for a causal effect to judging that the effect is real. Quant approaches (and, to be fair, some qual approaches) suggest that using their special method gives you a **free ticket** to make this leap. And indeed different methods include different ways to reduce different kinds of bias which mean you can be more confident in making the leap. But I'd say there are no free tickets. No way of an evaluator getting out of the responsibility of making the final evaluative judgement, however clever and appropriate your method.

(You could argue that FCM and Systems Dynamics do arithmetic on the strengths of connections. Perhaps that makes them quant methods.)

Seen this way, in essence qual and quant impact evaluation are not alternatives or competitors. They are different ways to do different things.

A second limitation of causal mapping is the difficulty it has in systematically capturing the strength or type of causal influence. It is relatively rare in open conversation for people to indicate in a consistent way the magnitude of the effect of C on E, or whether C was a necessary or sufficient condition for E or precisely how certain they are about the connection. There is of course scope for framing questions to encourage people to ascribe weights to their answers, which can then be incorporated into the way maps are constructed. But imposed precision risks turning

into spurious precision, and stronger framing of questions may distract from other issues and nuances that more open-ended questioning might otherwise have elicited.

Task 3: Answering evaluation questions

Causal maps help us to assemble evidence for the causal processes at work in specified domains, including the influence of activities being evaluated. They can also help expose differences between the evidence given by different sources and differences between the analysed data and theories of change derived from other sources, including those officially espoused by the commissioner of the evaluation (Powell et al., 2023). The identification of differences in understanding can then feed into further enquiry, analysis and action concerning why people have different views, what the implications of this are and how these might be addressed.

Focusing on causal claims is of course only one way of answering evaluation questions from a corpus of text data. But it is productive because many evaluation questions are at least partly about causation and causal contribution, and we have found that causal mapping points to possible answers to these questions relatively rapidly compared to more generic QDA approaches. Answering questions about efficiency, effectiveness, impact and sustainability, for example, all depend on identifying the causal effects of a specific intervention, be they perceived as positive or negative, intended or unintended (OECD, 2010). Even 'relevance' can have a causal interpretation in the sense that an intervention is relevant if it is doing the right thing: Whether it is likely to help to address the needs of stakeholders is at least partly a judgement about its causal powers.

For a data set comprising hundreds or thousands of links, an unfiltered global map of all the links is a bewildering and useless 'hairball' that includes everything but highlights nothing.

Causal mapping can help reconstruct a program theory empirically

To evaluate a program, the evaluator can use Contribution Analysis (CA) (Mayne, 2012). We start with a program logic or Theory of Change (ToC), consisting of possible pathways from interventions to outcomes, and collect existing or new evidence for each link. However evaluators can often not assume that the ToC underpinning a program aligns with the realities on the ground, or they may uncover outcomes not anticipated in the original program design - see Koleros & Mayne (2019). We have argued (Powell, Copestake, et al., 2023, p. 114) for a generalisation of CA in which evidence relevant to constructing a program theory, as well as evidence for the causal influences flowing through it, are both collected at the same time, without the evaluator (necessarily) having a prior theory. In this sense, following Mayne, "program theory" need not be something that any person necessarily possessed or articulated at the time, but is something which can be approximated and improved during the evaluation process.

(Re-)constructing program theory empirically in this way is an essentially open-ended, qualitative problem. Closed data collection methods are not suitable because we cannot measure what we do not yet know. Open-ended, qualitative methods to (re-)construct a theory are notoriously time-consuming and are usually heavily influenced by researcher positionality (Copestake et al., 2019).

Powell, Copestake, et al (2023, p. 108) present this task as gathering and synthesising evidence about "what influenced what", evidence which is simultaneously about theory or structure and contribution. Each piece of evidence may be of differing quality and reliability and about different sections of a longer pathway, or multiple interlocking pathways, and may come from different sources who see and value different things.

References

Mayne (2012). *Making Causal Claims.*

The result of an evaluation is a qualitative causal model

Evaluation outputs as a model

What is the output of an evaluation? We have a report, hopefully answering the evaluation questions. (In the sense of developmental evaluation perhaps some learning has taken place as well, or as a main output, but this is not what I want to address here.)

The final product can be more than a report: it can be a kind of model or knowledge graph which in principle can be queried to answer even new and unexpected questions.

[Causal maps are knowledge graphs, but with wings](#)