



QUALITY ASSURANCE IN CAUSAL MAPPING – ENSURING ROBUST AND RIGOROUS CONCLUSIONS AND INFERENCES

📅 25 Apr 2026

Quality assurance in causal mapping

From claims to conclusions

Causal mapping, as we practise it, is not a method of causal inference. We have always been clear about this. The fact that twenty people, or twenty thousand, claim that X influences Y does not on its own warrant the conclusion that X really does influence Y. The job of causal mapping is to assemble the claims so that an evaluator or researcher can make a judgement, not to make the judgement for them.

The longer argument for this conservative stance is in our companion paper [Minimalist coding for causal mapping](#) and in [Our approach clearly distinguishes evidence from facts and does not automatically warrant causal inferences](#); see also (Powell et al. 2024; Powell et al. 2023).

Heres' what we mean with our distinction between links (individual causal claims) and bundles of multiple such links between one cause and one effect: [Bundle of Links – definition](#).

Now that AI lets us scale a single project to tens or hundreds of thousands of causal claims, the gap between "we have many claims" and "we have warranted conclusions" matters more than ever. Practitioners need ways to cross this gap that are practical, transparent, and modest in their epistemic commitments. The Causal Map app helps at several distinct moments, which together we call quality assurance.

This paper covers the general picture. The bundle assessment phase, the most substantive single addition, has its own paper: [Assessing quality or robustness of](#)

[evidence for a causal link based on a bundle of coterminal causal claims.](#)

The same logic applies even when the links are not strictly causal: in social network analysis or other map-based work, you may still want to go from a mass of raw claims to a smaller set of checked or verified links. The mechanics described below work in the same way, though our main focus is specifically on *causal* links.

A note on "evidence"

We have been criticised for calling the mass of causal claims "evidence". The objection is fair: a claim is not really evidence until it has been weighed against something. Moving from a coded claim to a warranted conclusion is exactly the Rubicon this paper is about. When we use "evidence" loosely, we mean only the body of claims that the evaluator can take into account, not anything pre-warranted. We will try to be careful with how we use the word "evidence".

Four moments for quality assurance

It helps to distinguish four moments at which quality assurance can be done. They are not exclusive; most projects use several, or other overlapping approaches. We have always assumed that evaluators and researchers will be doing serious quality assurance when crossing the Rubicon from claims to conclusions, but this is the first time we have tried to address this task in a little more detail and point out how the Causal Map app can help.

(This is separate from the way causal inference is done specifically in the Qualitative Impact Protocol (QuIP) (Copestake et al. 2019) – although QuIP projects often use causal mapping and the Causal Map app, they have a more specialised and specific set of supports for causal inference.)

First, at coding time, individual claims can be tagged or qualified.

Second, the bundle assessment phase: a separate stage in which the analyst, optionally with AI assistance, judges each bundle of co-terminal claims and either creates for it a derived "assessed link" or declines to.

Third, at the level of often indirect (sets of) pathways e.g. from an intervention to an outcome.

Fourth, overlapping with the third, the vignette layer: the AI vignette feature drafts a commentary on a chosen view that helps support inference by drawing on the underlying quotes and source metadata.

First moment: Coding-time qualifications

The first step to quality assurance of a claim is to tag it. The Causal Map app has always allowed free-form tags at the link level. A tag like `#doubtful` records a misgiving while coding without slowing things down. Later, you can filter such links in or out. Tags are freeform: you can create `unclear` or `#decisive` or anything you want.

Beyond tags, you can add custom **columns** to your links table. Here are two common columns.

A **conviction** column records how sure the source sounds about the claim. In practice most claims are unmarked: people just say "X influenced Y" without qualification. A workable three-point scale is weak / neutral / strong, with perhaps ten per cent at each end and the bulk in the middle. This is not a coding of the causal strength of the link itself but only a coding of how confident the source sounds.

If desired, a **strength** column captures cases where a source explicitly says the influence is strong or weak. Our experience says that humans rarely actually mention this in speaking and writing: again, the bulk of claims is likely to be assessed as neutral: no explicit information about strength. But it might be useful to record strength, for example because we might want to filter out claims about weak strength, or examine only the strong ones.

We suggest caution in interpreting these kinds of scale as ordinal (small, medium, large; or 1, 2, 3). Linguistically, at least these two cases rest on the idea that the default claim is unmarked or neutral, which is not the same as "middling". There was simply no need to mention or even think about this aspect. The fact that most people do not mention the strength of a causal link when talking about it does not mean they think the links were of "medium" strength. It just means it did not occur to them to think about or

mention the strength, or that the idea of strength is not even useful or applicable in this case.

For more background on why we have been reluctant to code strength: [Our approach is minimalist — we do not code the strength of a link](#)).

Other columns are possible. Tom Aston and Marina Apgar propose around ten possible link-level or bundle-level fields in their work on rubrics (Aston 2019; Mayne 2014). The framework is open: you decide what matters in your project for supporting the conclusions you want to make.

The Causal Map app now supports creating custom link columns like this either before coding or even on the fly, in the middle of coding.

You can also add custom columns for **sources** rather than links, for example distinguishing reliable from unreliable sources, or recording role and position. Because every link belongs to a source, these scores become available at the link level via a join, and you can filter accordingly.

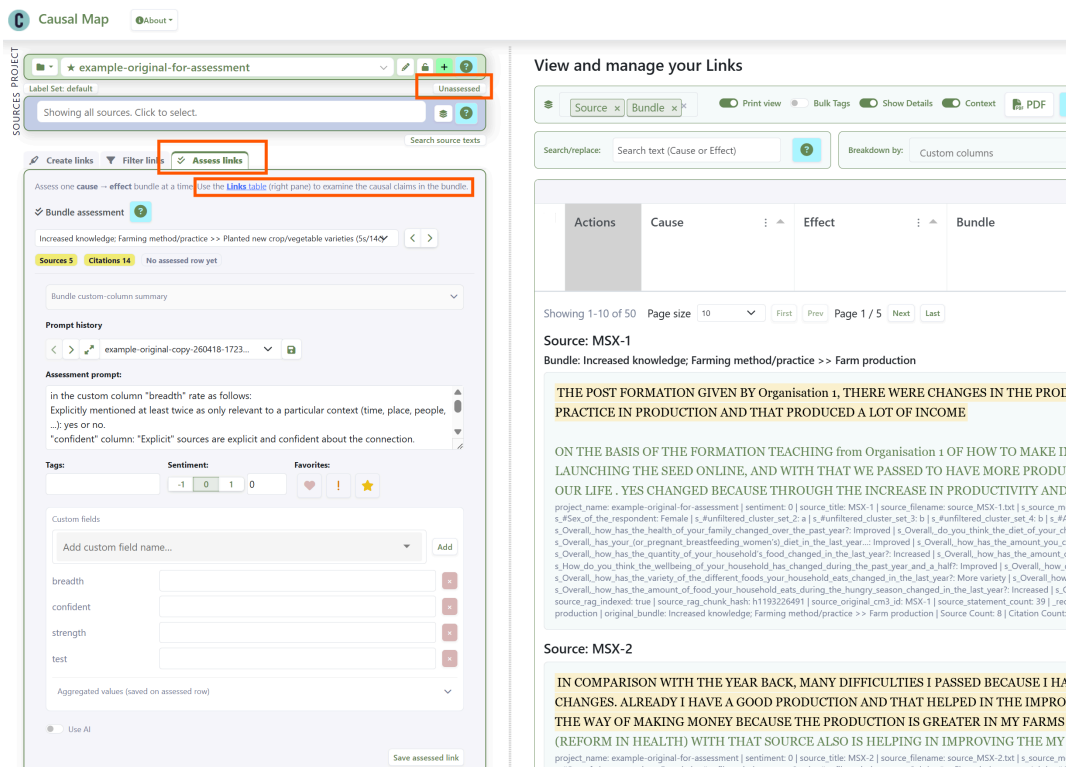
Bundle-level summaries make this useful. When you look at a bundle of claims for X influences Y, the app summarises the distribution in a sub-panel of the Assessment panel, for example reporting that in eighty per cent of cases conviction was neutral; in fifteen per cent the source emphasised they were sure. This is helpful both as a backdrop for human judgement and as a filter (e.g. exclude links where the source said they were uncertain). See [Coding with and using link metadata](#) for the mechanics.

Second moment: The bundle assessment phase

This warrants its own paper; see [Assessing quality or robustness of evidence for a causal link based on a bundle of coterminal causal claims](#) for the detail. In outline:

Once coding is finished and any cleaning has been done, you fix on a set of bundles you want to take seriously. These are the bundles that survive your filters, perhaps after zooming to a higher level of the coding hierarchy and restricting to particular sources or subgroups. There might be five or fifty or a hundred such bundles of links. This is the data you are going to base the rest of your analysis on.

You then look at each bundle, with all its underlying quotes and source metadata, and decide whether the body of claims is enough to vouch for a second-level "assessed link" between the two factors. The assessed link is a new type of object in the links database. By default it inherits the citation count and source count of the underlying bundle and can carry additional scores from custom columns. Some bundles will not produce an assessed link at all, because you have judged the evidence too thin. You simply skip the bundle without creating any assessed link. Or you create a link with a custom column "Passed?", with value = "Fail".



Creating new

individual "assessed links" from bundles of links, bundle by bundle, in the Causal Map app

You can paging through the bundles by hand, or you can let the AI do a first pass against a rubric you supply, and then review. The app will not let you create assessed links — either manually or with AI — until you have written your criteria into a rubric or prompt sub-panel. This is on purpose.

The rubric might be a five-level scale like the one Jewlya Lynn and colleagues used in their fishing industry retrospective (Lynn 2025), or just yes/no. Or you might want to

create multiple dimensions like "confidence" and "degree of triangulation". The decision is yours.

The result of this Bundle Assessment process is a parallel map. The unassessed claims remain in the database, but a switch in the app lets you view only the assessed links. A typical project might go from 1000 raw claims to 500 filtered claims in 30 bundles to 25 assessed links. You can use the newish "Map Custom Columns" filter to apply custom formatting to your links in the final maps, by source count, citation count, or any custom score (degree of triangulation, for example).

For writing up, the assessed map gives you a cleaner basis for argument than the raw claims, while preserving full provenance underneath.

Third moment: Pathways and the transitivity trap

Even when each link, or each assessed link, is now well grounded, your work is not finished.

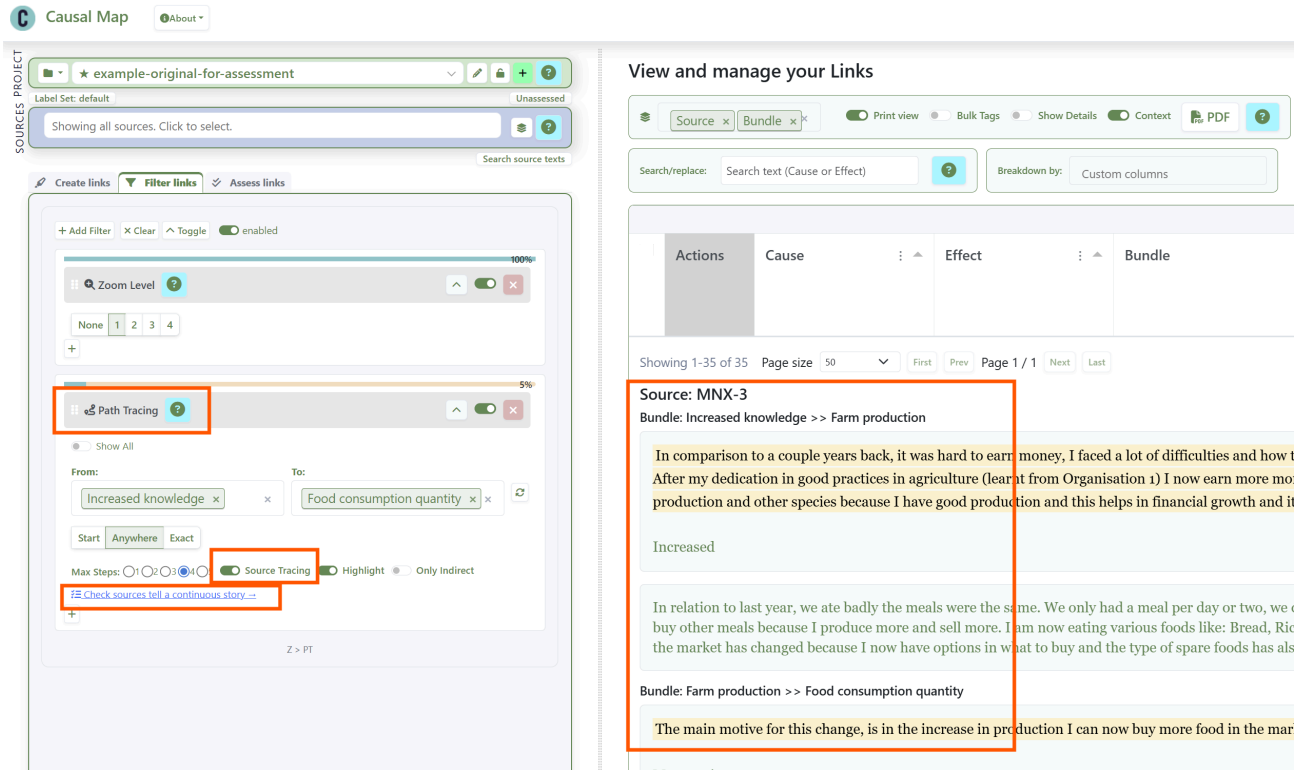
Often you will need to draw conclusions not just about single influences of B on C but about a whole overlapping network of mostly indirect links from B to C via E, F, G and so on.

Two app features help.

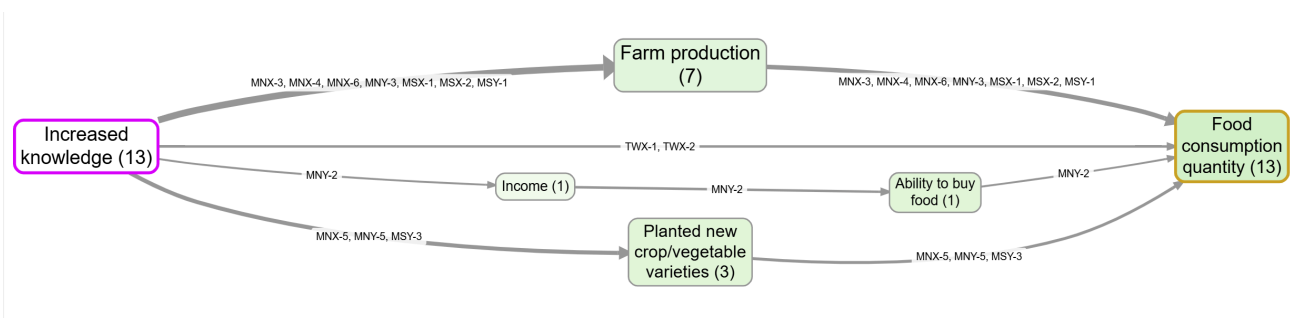
First, filters the map to the links that lie on some pathway from your chosen start factor to your chosen end factor, within a set number of steps. It excludes all links which are not on such a path, to make it easier to examine the evidence for whatever conclusion you want to draw.

However, from "A influenced B" and "B influenced C" you cannot in general conclude "A influenced C", because the contexts in which each step holds may not overlap. This is [The transitivity trap](#), the single most important challenge for any approach that uses directed network diagrams. So Causal Map provides Source Tracing as the stricter version of Path Tracing: it finds only sources which have any pathways all the way from A to C and keeps only those pathways, and then combines all such pathways into one map. This is the conservative move when you want to avoid stitching fragments of different stories together. Every link is then part of at least one complete story from one

source from A to C. A new button in the app opens the links panel arranged in such a way that you can review the evidence source by source and judge whether each respondent's account is internally coherent.



Setting up Source Tracing from Increased Knowledge to Food Consumption Quantity, and examining the corresponding narratives.



The corresponding map, showing source IDs and source counts.

If you have already run a bundle assessment, there is a choice to make: source-trace on the assessed links or on the unassessed ones? With the assessed links you get clean source and citation counts but no direct view of the quotes. With the unassessed links you get the quotes but a busier map. In practice you may want both, in different views.

Fourth moment: Vignettes

Some readers want to see the map. Others want to be told what it says. The [vignette](#) feature drafts a narrative commentary on the current map view, drawing on factor labels, link metadata, source-level metadata, and the original quotes. You write the prompt, so you control the focus. A common use is to ask for a commentary on the pathways from an intervention to a chosen outcome from the perspective of individual sources, discussing how coherent each source's story is. The AI is doing nothing more than a careful reader could do given the same inputs, and the patience to examine the quotes behind each link.

Vignettes can be created with the specific task of answering quality assurance questions like: is each link really part of a coherent complete and consistent story from source factor (e.g., Intervention) to target factor (e.g., Outcome)?

Vignettes

Model:
Gemini 2.5 Pro

AI processing uses the project's AI Processing Region (Project → Edit project).

Enable checking (second AI pass)

generic-source-tracing - 25/04/26, 18:36

Whole-map prompt:
What causes what (cause → effect) along with the quote where it was mentioned.

Don't describe the map as such, only its contents. Use the existing cause/effect labels where possible.

Each individual link in this map is supposed to be part of a complete story from "Increased Knowledge" all the way to "Food consumption quantity" as told by a single source.

Your main task is to report how much is this the case: do the quotes for each source really represent such a coherent story? Provide just a few key quotes (and cite the source ids of the quotes) which best

[Write Whole-map Vignette](#)

Vignette generated

Based on the data, the quotes provided by individual sources largely represent coherent causal stories explaining the journey from connecting multiple steps of a causal chain within their testimony.

Knowledge leading to Production and Consumption

A primary narrative described by several sources is that increased knowledge, gained from an organization, led to improved farm logically. For example, one participant explained how lessons improved their output, and then linked that output directly to what

In comparison to last year where things were very difficult because I had challenges in earning money, I worked for some farm the lessons from Organisation 1 I now produce a lot and I have a good income from my products. I now produce and sell and grow well.
— MNX-6

This same source then connected this increased production to their household's consumption.

In contrast to last year we barely ate, our meals were salt and water, we only had a meal or two and we never varied our meal am eating more as I produce more and I sell more as well. thing like: Bread fish, rice, potato, cassava, spaghetti, and tea. Yes c store food, I chaff the goods and place into bags which we keep at home until this season come again, I barely sell goods as as well because I buy other kinds of seeds for my garden. The reason for the change (increased food consumption) is in the v
— MNX-6

Another source also provided a clear, two-part story, first linking knowledge to production and then production to consumption.

An automated Vignette for the same map, tasked with examining whether the evidence for each pathway is coherent.

What the app does not do

At no point does the Causal Map app move on its own from claims to facts. Causal mapping as we see it is still, on its own, not a method of causal inference but more of a way to *identify and organise the evidence* in order for the evaluator or researcher to make causal inferences, especially when assisting established methods like Contribution Analysis or QuIP. Still, in the past we have perhaps not done enough to say how exactly to do this or to make it easier to do. This post hopes to redress that.

The warranting is always the evaluator's. We provide structures (tags, columns, the assessed-link switch, source tracing, vignettes) that make warranting easier, more transparent, and more auditable. We do not provide an engine that turns "twenty people said so" into "therefore it is so".

The opposite design, in which an algorithm rules on causal truth from coded text, would either smuggle in strong assumptions about variables and functional forms (which we argue against in [Our approach is minimalist – we do not code the strength of a link](#) and at length in our minimalist coding paper) or quietly conflate evidence volume with effect size, which Causal Map has always been at pains to avoid. As we put it elsewhere, "a coded link is first and foremost 'there is evidence that a source claims X influenced Y', not a system model with weights or effect sizes" (Powell et al. 2024).

A typical workflow

A causal mapping project that uses these features looks roughly like this. You code a corpus, in vivo, and end up with several hundred or several thousand raw claims, each with a quote and a source. You tag occasional claims as doubtful, code conviction where it stands out, and code source reliability in the source metadata. You filter to a maximal set of bundles that matter for your evaluation question, perhaps zoomed to a level of abstraction at which your factors are useful. You run the bundle assessment phase, by hand or with AI assistance plus review, against a rubric you have written down. You arrive at a much smaller set of assessed links, each of which you are willing to vouch for. You trace pathways, source by source where it matters, between interventions and outcomes. You may ask for vignettes that walk a reader through what the assessed map says.

None of this is causal inference in a strict sense. It is a disciplined way to assemble evidence, weigh it transparently, and reach conclusions that you can defend.

This all works, we use it every day in our consultancy work at Causal Map Ltd., but it is still also a work in progress, so if you are interested or have something to say, do get in touch.

References

Aston (2019). *Contribution Rubrics*.

https://media.licdn.com/dms/document/media/C4D1FAQE1laRiOvrFrQ/feedshare-document-pdf-analyzed/o/1620553059516?e=1687996800&v=beta&t=bFr7dhpZ-slluV8ne1cERFelwINIaEzsQN8fiF_75gQ.

Copestake, Morsink, & Remnant (2019). *Attributing Development Impact: The Qualitative Impact Protocol Case Book*. March 21, Online.

Lynn (2025). *HU Seafood Retrospective*.

<https://www.policysolve.com/resources/retrospective>.

Mayne (2014). *Contribution Rubrics 1*.

Powell, Larquemin, Copestake, Remnant, & Avard (2023). *Does Our Theory Match Your Theory? Theories of Change and Causal Maps in Ghana*. In *Strategic Thinking, Design and the Theory of Change. A Framework for Designing Impactful and Transformational Social Interventions*.

Powell, Copestake, & Remnant (2024). *Causal Mapping for Evaluators*.

<https://doi.org/10.1177/13563890231196601>.