# 🌻 Causal mapping for evaluators – Summary (eval2024)

(Powell et al., 2024)

Source: (DOI: 10.1177/13563890231196601)

- **History / lineage (why this isn't "new", just under-used in evaluation)**: Causal mapping (diagramming "what causes what" using directed links between factors) has been used since the **1970s** across disciplines (e.g. Axelrod-style **document coding** of causal assertions; management/OR traditions emphasising maps for **decision support**; comparative methods like Laukkanen's work on **standardising factor vocabularies** and combining maps). The evaluation literature has relatively sparse, inconsistent "causal mapping" usage; this paper synthesises the wider literature and re-specifies it for evaluators.

- **How we pitch it to evaluators (the niche)**: treat causal mapping as a **discrete evaluation task**: (i) systematically **assemble causal evidence from narrative sources** into an explicit link database with provenance, then (ii) separately use that assembled evidence to make evaluative judgements about "what is really happening". This is positioned as a way to work with large bodies of **messy, heterogeneous** qualitative causal data (different boundaries, contexts, specificity, and ambiguity) without forcing early convergence on a single prior ToC.

- **How causal mapping differs from adjacent approaches**:

- **Primary object is evidence-with-provenance**: causal mapping is explicitly about *who/what source said what link*, not a modeller's best estimate of system structure.
- **Epistemic first, ontic later**: unlike approaches mainly aimed at simulation/prediction (e.g. SD/BBNs/CLDs/FCMs as typically used), causal mapping foregrounds **organising claims/evidence**; inference about reality is a later step.
- **Lightweight causal typing**: it usually does not require consistent weights/functional forms/necessity-sufficiency labels; it can incorporate them when elicited, but warns about spurious precision.

- **How causal-mapping approaches differ among themselves (key axes)**:

 

- **Mode of construction**: coding **documents** vs coding **interviews** vs **group** map-building (consensus/problem-structuring) vs hybrids.
- **Elicitation openness**: **closed** (pre-specified factor lists) vs **open** (respondent-generated factors), with chaining variants (forward/back).
- **Single-source vs multi-source & context handling**: idiographic maps vs aggregated multi-source maps; whether and how you track **case/context metadata** to avoid invalid transitive inferences.
- **Coding philosophy**: "factors as variables" vs "factors as **changes**" (e.g. QuIP-style); whether polarity/opposites are represented as separate factors/links or handled differently; extent of factor-name **standardisation/merging/nesting**.

- **Problem / motivation**: Evaluators need to represent (a) what causally influences what **in the world**, and (b) what different stakeholders **claim/believe** causally influences what. Causal mapping—defined as the **collection, coding, and visualisation of interconnected causal claims** with explicit **provenance**—is widely used outside evaluation, but under-specified in evaluation practice/literature.

- **Core argument (the "Janus" dilemma + resolution)**:

- **Janus dilemma**: Causal mapping faces two directions—maps can be read as **models of beliefs** or as **models of causal reality**; in practice these get blurred unless source information and analysis steps are explicit.
- **Resolution**: Treat causal maps primarily as **repositories of causal evidence** (epistemic objects), not as direct models of either beliefs or reality. Maps then support structured questions like: *Is there evidence X influences Z? Directly/indirectly? How much evidence? How many sources? How reliable?* The *evaluation* step that judges "what is really happening" is distinct and subsequent.

- **What causal maps encode (and don't)**:

- **Epistemic content**: Map elements are claims/perceptions/evidence, not facts.
- **Causal semantics are usually coarse**: ordinary language claims typically encode **partial influences**, not total/necessary/sufficient causation; coding a link need not assert evidence quality (though you may later weight/filter by quality).
- **Multiple sources + contexts**: maps may be single-source or multi-source; inference across sources requires care about **which case/context** each link refers to.
- **Boundaries are often messy**: system boundaries are frequently loose/implicit; mapping can proceed, but ambiguity must be managed rather than hidden.

- **Causal mapping in evaluation = 3 tasks (workflow)**:

- **Task 1 — Gather narrative causal material**: interviews, open-ended survey questions, document/literature review, archives/secondary text, or consensus processes (e.g., Delphi, participatory systems mapping). Elicitation may use **back-chaining** ("what influenced X?") and **forward-chaining** ("what followed/could follow?"). Question framing affects factor semantics (e.g., QuIP tends to elicit **changes** like "reduced hunger" rather than variables).
- **Task 2 — Code causal claims ("causal QDA")**: unlike standard thematic QDA (codes = concepts), causal QDA codes **links**: each highlighted quote yields an **influence factor → consequence factor** pair; factors mainly exist as endpoints of links. Labelling can be **exploratory/inductive** (curate a common vocabulary across sources) or **confirmatory** (codebook from a ToC/prior work), with sequencing cautions to reduce framing/bias. Manual coding is costly; partial automation via NLP/ML is possible but not the focus.
- **Task 3 — Answer evaluation questions using the link database**: global maps become "hairballs", so analysis should generate **selective maps** aligned to questions (e.g., consequences of an intervention; causes of a valued outcome). Techniques include bundling **co-terminal links** (thickness/count), producing frequency-based overview maps (caution: rare-but-important links), rolling-up hierarchical factor taxonomies (with caveats), and limited quantitative summaries (warning: sensitive to coding granularity).

- **Limits / risks**:

- **Inference depends on source credibility**: stronger conclusions require explicit, context-specific **rules of inference** (e.g., independent mentions threshold + theoretical plausibility + bias-mitigation steps).
- **Effect strength/type is hard to capture**: respondents rarely provide consistent magnitudes/necessity/sufficiency/certainty; forcing weights risks **spurious precision**.
- **Transitivity is both payoff and trap**: inferring (C \rightarrow E) from (C \rightarrow D) and (D \rightarrow E) is powerful for indirect effects, but can be invalid when links come from **non-overlapping contexts**; valid inference requires attention to the **intersection of contexts**.

- **Concrete analytic contributions highlighted**:

- Treat diagrams as an **index into the underlying corpus**: tool support should allow tracing from any link/factor back to transcript excerpts + source metadata.
- Quantify robustness of evidence-based "arguments" along paths using **maximum flow / minimum cut** on the causal-claim network (how many claims would need removal to eliminate all paths between (C) and (E)), plus **source thread count** (how many distinct sources each provide a complete path).

- **Conclusion / evaluator-facing payoff**:

- Helps evaluators (i) assemble narrative evidence about intervention and contextual influences (direct/indirect, intended/unintended), (ii) search/summarise/select quotations systematically, (iii) increase transparency/peer-reviewability of qualitative causal reasoning, (iv) communicate complexity with readable graphics.
- Key discipline is a **two-step separation**: first assemble and organise causal evidence; then judge what is actually happening—avoiding premature constraint of data collection to fit a prior ToC that stakeholders may not share.

## References

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

 