Bath Social & Development Research Ltd

Qualitative Impact Assessment Protocol

From narrative text to causal maps: QuIP analysis and visualisation

October 2021

1. Introduction

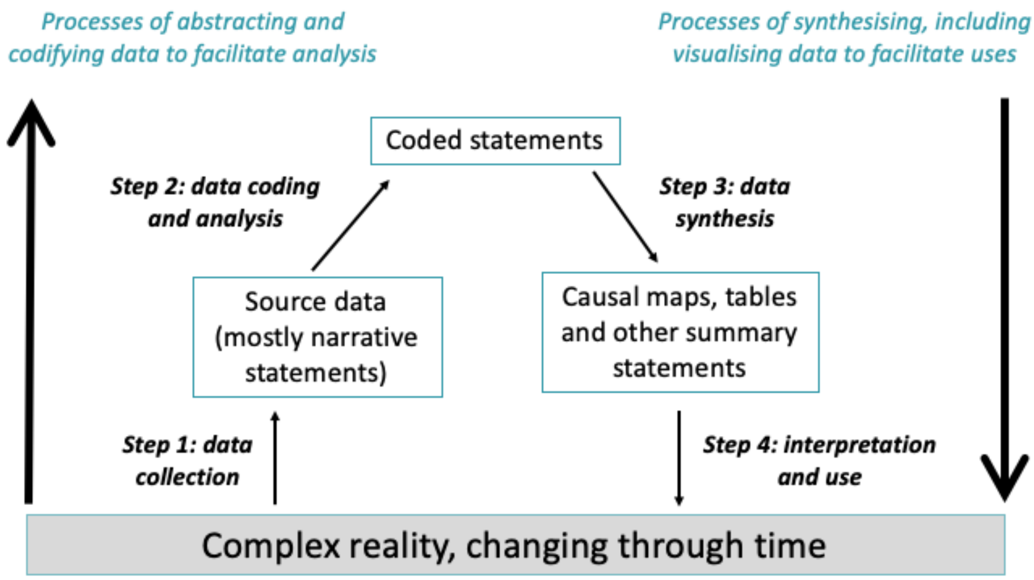
What sort of data do QuIP studies produce? This note explains the link from written reports of what people have said to production of useful summaries or distillations of this material. We cannot ignore entirely how the source data was collected in the first place, nor the final purpose of the whole exercise. But the focus here is firmly on the intermediate analytical step – i.e. how the raw data is summarised, sorted, simplified, synthesised or otherwise presented (literally 're' presented) in order to make it more useful.

At its simplest, QuIP analysis entails picking out short statements containing causal claims (also referred to as 'stories of change') attaching codes to them, and sorting these blocks of coded text in ways that make it possible to generalise from them about the causal processes being reported. More specifically, QuIP analysis produces useful visual summaries of what the data contains in the form of causal maps. It also does so in a flexible and transparent way that can be audited and replicated.

Much of this note comprises 'how to' descriptions of selecting, coding, filtering and using data to produce causal maps. More comprehensive 'how to' guidelines for coding QuIP data and using the *Causal Map* software are available elsewhere. Here we are more concerned with 'why' questions, and how the outputs of QuIP analysis can complement other forms of enquiry.

The scope of this note is further clarified with reference to Figure 1 below. This depicts QuIP analysis as a set of actions that occur through time (moving from left to right), while also involving *abstraction* from reality followed by synthesis and reengagement with reality (moving up and down). The diagram also breaks this process into four steps. First, we collect and record perceptions of what is going on in the real world, mostly in the form of raw narrative text. Second, we sort this data and attach codes to it in order to facilitate its analysis. Third, we construct diagrams and other summaries of the data to facilitate useful generalisation. Fourth – and this is the acid test – we draw on the analysed data to re-engage with reality aided, hopefully, with useful additional insights into what is happening and why.

Figure 1. Summary visualisation of the overall QuIP research process



QuIP analysis is messier in practice than this model suggests. The analyst must hold all the steps in mind at once, and progresses through them iteratively, with small feedback loops between the steps (not shown). For example, construction of causal maps (Step 3) is likely to prompt revisions to initial sorting and coding (Step 2). In addition, summaries of the data do not solely rely on coding. For example, an analyst may also draw on experience and intuition to pick out key points and illustrative material from the text.

QuIP data collection (Step 1) has distinctive features that reflect the focus on eliciting causal stories. For example, interviews are structured around discussion of changes in perceived outcome domains, and then 'backchain' by asking questions about potential drivers of those changes. However, this paper is concerned mostly by subsequent steps. The next section goes into much more detail about Steps 2 and 3, particularly how coding aids the construction of summary causal maps from the raw narrative data. The note then reflects on how this whole process relates to the familiar but often confusing distinction drawn between qualitative, quantitative and mixed methods of enquiry. This in turn enables us to make some concluding observations about how the QuIP compares with other approaches (quantitative and qualitative) to building evidence of causal processes, and how it can usefully be combined with them.

2. QuIP analysis: from coding narrative text to construction of causal maps

A typical QuIP comprises narrative text extracted from interviews with 24 individuals and four focus groups – let's refer to these as 28 data sources. This is loosely organised by predetermined outcome domains – let's say for simplicity there are ten of these – with each source asked to identify what changes they perceive to have taken place in each within a carefully specified period of time. It follows that we face a problem of how to capture and convey what is most useful about the activity the commissioner is interested in across 280 discrete sets of textual data. Within each set we are looking particularly for causal claims that link stated outcomes in each domain (such as a change in household food consumption) to one or more drivers of change.

Before coding causal links, it is useful to clarify what respondents perceived to have been the overall change in each outcome domain. To do this typical QuIP interviews (but not focus groups) include a small number of domain specific closed questions. These ask respondents to indicate whether change in the selected outcome domain over the specified period has been positive or negative for them. Reports typically start by presenting this data, using a table like Figure 2. This provides a rapid visual indication of whether the story of change they are collectively telling is broadly positive, negative or mixed. It may also be possible to pick out patterns in the responses – e.g. according to respondents' wealth, age, gender or place of residence. One purpose of this table is to gain an overview of respondents' experience of change in a way that highlights variation across all sources, without hiding individual perceptions behind aggregate statistics. It also makes sense to summarise changes in outcome domains (if any) before reporting on what respondents claim is causing them. The table also whets the appetite, ahead of presentation of the main dish – causes of change, collected through open-ended questioning.

Moving on now to preparation of the main dish, the first step is to identify, classify and code different sorts of causal claims (e.g., 'X caused Y', or 'Y happened because of X and Z') that link outcomes back to what respondents perceive to be their main causal drivers. QuIP coding differs from more generic or thematic coding of qualitative data because codes or tags are not linked to single concepts but to causal statements, which have a minimum of two elements:

- **a driver of change, influence** – i.e. the reason given to explain any change or outcome;
- **an outcome** or **consequence** – i.e. what the specified driver of change caused to happen.

Figure 2: Example - Closed question responses

	Wealth Group	Home food production	Money from livestock	Money from other sources	Quantity of food	Variety of diet	Purchasing power	Value of assets	Health of children	School attendance	Amount children working
DHFC-2	Middle	-	-	-	-	-	-	-	=	+	=
DHMC-4	Middle	-	-	-	-	-	-	-	=	=	=
DHMC-5	Middle	=	-	+	+	+	+	-	+	+	+
DHMC-6	Middle	+	+	+	+	+	+	+	=	=	=
DHFC-7	Middle	+	+	-	+	-	-	-	+	+	+
DHMC-11	Middle	=	=	+	+	=	-	+	+	+	-
[illegible]	Middle	=	=	-	-	-	-	-	+	+	+

Note: columns show outcome domains, rows represent individual respondents or cases. The +, - and = signs indicate whether respondents' said overall change in the specified domain over an agreed period was positive, negative or neither.

Analysts typically choose codes for factors inductively, to reflect what they find in the statements. At the same time, they can augment codes with additional bits of information to assist subsequent analysis. More specifically, most QuIP studies use three additional coding options.

- First, analysts can tag drivers that they recognises as linked to a specific intervention, whether explicitly or implicitly, e.g. 'nutrition education from the government village health worker [E]', where E indicates an explicit reference to the intervention being evaluated. This option facilitates aggregation of evidence about the effect of a particular intervention.
- Second, tags can be used to indicate whether the analyst interprets a respondent as associating an outcome with either a positive or negative sentiment, e.g. 'Children attending school more [P]', where P indicates what is interpreted as a positive outcome. This permits searches of all positive or negative outcomes.
- Third, codes can be 'nested' hierarchically as belonging to a wider or parent theme, such as Health, e.g. 'Health; Children ill less often', where 'children ill less often' is an example of a factor relating to 'Health' overall. This would enable analysts to simplify or zoom into maps at different hierarchical levels depending how much granular detail they want to see.

While it is possible to adapt Excel and other software packages to do this kind of coding, experience in analysing data in this way to produce causal maps as well as tables led us to collaborate in the creation of a bespoke software package - www.causalmap.app. This is the package we now use and recommend for all QuIP analysis, and in which the illustrative material below has also been produced.

Once the narrative statements are coded, the analyst can start aggregating causal maps across the whole dataset - linking higher-level outcomes back to intermediate influences and underlying drivers of change. A causal map that reveals all coded causal statements is often very complicated and hard to interpret, hence part of the role of the analyst is to identify which maps to draw and to share. A causal map can use data from all or a selection of sources, and can be filtered by types of factor labels or frequency of links or factors. Choice of what maps to draw may be guided by the commissioner of the study, or it can be more exploratory. Either way, it should always be possible for a separate analyst to replicate the same causal map, so long as they are using the same coded database and are furnished with details of what filtering and other decisions the first analyst chose to make.

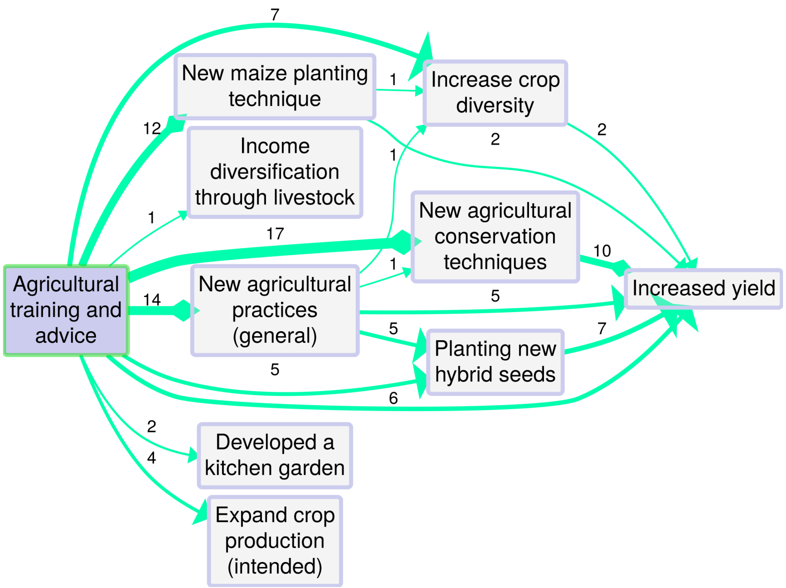
Typical questions that QuIP causal maps address include the following:

1. Is there evidence that Intervention X is having the expected effect on intended beneficiaries, and if so, *how much* evidence is there?
2. Did other factors affect expected outcomes, and if so, how much evidence is there?

3. Has the programme had any unanticipated effects, positive or negative?
4. What drivers of change or patterns can be identified that could inform future programme design?
5. Are there significant differences between the maps derived from different kinds of respondents (e.g. by wealth category, location, gender, age groups etc)?
6. How do the causal maps drawn from the narrative statements of different stakeholders compare with the theory of change already held by the commissioning organisation?

Figure 3 shows a map from the application Causal Map, showing coded causal statements for a project that provided farmers with agricultural training and advice in order to increase crop yields. The map has been filtered to show only outcomes downstream of the influence factor 'Agricultural training and advice'. Numbers shown indicate how many times the links were made across all interviews.

Figure 3. Example causal map looking at outcomes linked to one driver :



This quantification of causal connections can also be presented in tabular form, breaking down how different factor labels have been used across all the data, or to compare different groups of respondents. The table below presents counts of the use of factor labels depending on their position as an influence (from) or consequence (to) factor. Of course, this does not present them in the context of the links that are made between them, which can be seen in the maps, but it serves as another perspective on the data and helps the analyst to know which key factors would be useful to start querying in the maps.

Figure 4. Example of a table listing factor counts

Factor frequencies

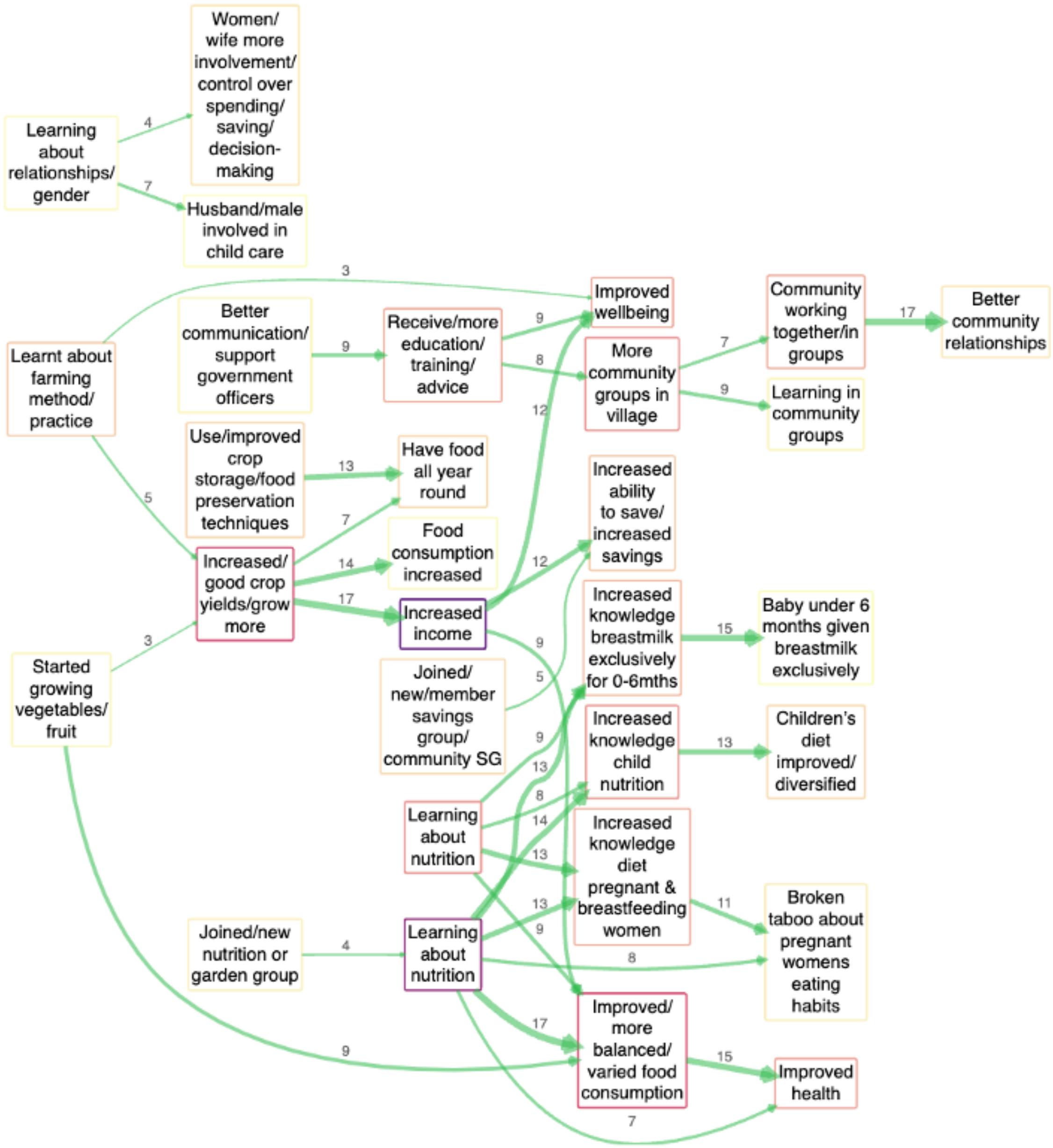
Showing numbers of links from and to each factor; factors are listed with most frequent first.

factor	from	to	total
(IEA) Increased purchasing power [P]	247	197	444
(IEA) Social Cash Transfer (OrgX) [E]	229	0	229
(HN) Increased food security [P]	84	123	207
(BF) Started, expanded or invested in business [P]	44	64	108
(IEA) Increased savings/loans [P]	85	6	91
(IEA) Increased income [P]	21	58	79
(IEA) Increased assets [P]	5	65	70
(BF) Agricultural training and advice [E]	68	0	68
(BF) Increased yield [P]	19	44	63

There are many ways to filter maps. For example Figure 5 has been simplified to show only the 30 most frequently cited factors and the 35 most frequently coded links. The numbers indicates the citation count – i.e. the number of times a link was coded. A red link means this has a negative impact on the linked outcome.

With the choice to select data for visualisation comes responsibility. The analyst can choose to simplify a map based on restricting the minimum number of factors or links, for example selecting only links cited more than 10 times, or the 15 most commonly used factors. This is a good way to convey the 'big stories' from a study, but it could also hide some of the more granular complexity and some interesting but less commonly cited stories. Similarly maps can be filtered by selecting only a subgroup of respondents, by showing only the links forward from a specific driver, or only the links back from a specific outcome. . To ensure they are not misread all maps should state clearly what selections have been made and why. For example, a map that is generated only by selecting forward causal links from a known 'project' activities (such as Figure 3) conveys an impression that the project is more important relative to other causal drivers than a map that shows all factors. For this reason, we are also cautious about implementing such choices algorithmically. But as experience with causal mapping grows it may be possible to develop some general rules or a 'grammar' for good map construction.

Figure 5. Example causal map simplified to show most frequent links and factors across a dataset



3. Being clear on numbers

The numbers used in causal maps and tables need to be carefully explained to readers to ensure it is clear what they mean in the context of a study. Readers should always be aware of the total number of sources used in a study (a source can be an individual interview, a focus group, a report etc.), and the type of count used should be clearly labelled.

Counts can apply to:

- Factors: these are the individual labels given to each end of a link
- Links: these are the links made between two factors

Most maps and tables will display citation counts - how many times a factor or a link has been used across all the statements in the dataset - including when repeated in different questions within one respondent's interview. This can therefore be higher than the number of sources

However it is also possible and important to check source counts - the number of sources where a factor or link is used. This is of course limited to a maximum of the total sources used in the dataset (whether this is interviews or reports).

These two counts tell us two different types of information, so it is useful to toggle between them. The citation count is simply a frequency count, counting every time a link or factor was mentioned and could be a very high number if it is mentioned often by many people. The source count shows how many people or groups, or reports, mentioned a link or factor. This is a useful distinction for the analyst who should be comparing these counts.

Consider the difference between the number of sources (out of 24) that made a specific causal claim, and the total number of causal claims coded. For example, our analyst might identify and count 24 statements that explicitly linked the project to an outcome. A reader's interpretation of this evidence might vary a lot depending on whether this 'citation count' of 24 statements came from just four respondents (hence repeated on average across six different questions within an interview), or from all 24 respondents and with reference to just one question. The quality of this indicator will also depend upon whether the 24 sources included any focus groups or just individual respondents.

There are other ways of counting the strength of links between certain factors. If using the application Causal Map, there is a built in algorithm which calculates the Robustness of an argument that one thing leads to another; how much evidence there is for the causal path(s) from selected influence factors selected consequence factors[1]. This helps when you want to compare pathways between different factors, and can't easily do this simply by looking at a map. It is difficult to keep track of the full pathways between factors if they involve multiple intervening factors. This calculation identifies how many pieces of evidence (individual mentions of individual links) would have to be deleted before the pathway collapsed. A larger number means more evidence; more pieces would have to be deleted for the hypothesis to be rendered false.

[1] See https://guide.causalmap.app/quantifying-causal-evidence.html?q=robustness#summary-2 for more

Figure 6. Example Robustness of Argument calculation

source	target	evidence
Paths from all sources to all targets		
(IEA) Org X Social cash transfer [E]	All targets	191
Paths from all sources to each target		
(IEA) Org X Social cash transfer [E]	(HN) Increased food security [P]	101
(IEA) Org X Social cash transfer [E]	(IEA) Increased purchasing power [P]	177

This number is shown in the "evidence" column in this example, where a large number means that there is lots of evidence for this pathway, not that the effect of the source is strong. On its own this number is of limited use, but when compared to other robustness queries for different pathways in the same dataset, the comparison can be useful.

Another reason for needing to make comparisons between absolute citation counts carefully is that they partly reflect the total number of respondents interviewed. For example, in a study based on 24 respondents, 6 out of 18 women interviewed might link participation in a training programme to improved wellbeing, while only 3 out of 6 of the men interviewed did so. The absolute figures indicated stronger support for the link from women (6 compared to 3), but the relative figures suggest otherwise, as only a third of women mentioned the link, while half the men did so.

Counting and visualising causal connections is an attempt to make dense narrative text much more accessible and comparable, but it also selects from and simplifies the underlying narrative data. However, the rich underlying text is not lost, and indeed one purpose of the mapping may be to draw attention to particular causal statements, stories or sources that are worth exploring and describing in more depth. For this reason causal maps are often accompanied by discussion and by reference to specific quotations. Each count of a causal claim reflects something that is qualitatively different and eyeballing the headline numbers needs to go alongside reading selected quotations: some because they sum up something repeated in different ways by several respondents, and others because the analyst regards them as particularly insightful or interesting in their own right.

In QuIP analyses, respondent voices are always front and centre: all coded causal connections link transparently back to the original text, so that anyone asking, "where did that link come from?" can read the researcher's record of the respondent's original words. One aim of QuIP reporting is to encourage readers to get involved with respondents' original statements and read them in context.

Figure 7. Example of how the quotes linked to specific links within a causal map can be displayed

From: (HN) Improved hygiene practices [P], To: (E) Reduction in mosquito breeding environments [P], 1 mentions

Such practices include use of the pit latrine which is covered after use, washing hands after visiting the toilet and before eating or handling any food, washing hands after changing baby nappies and keeping the household and its surroundings clean to avoid the breeding of mosquitoes which cause malaria [Source: mjf-7]

From: (HN) Increased WASH knowledge, To: (HN) Improved hygiene practices [P], 1 mentions

They have adopted hygienic practices at the household to avoid contamination which may lead to disease outbreak. Organisation X has been promoting hygienic practices among the beneficiaries of their interventions since last year. Such practices include use of the pit latrine which is covered after use, washing hands after visiting the toilet and before eating or handling any food, washing hands after changing baby nappies and keeping the household and its surroundings clean to avoid the breeding of mosquitoes which cause malaria. Before Organisation X introduced their interventions, the family used to access information regarding sanitation and hygiene from the Health Surveillance Assistant who is a government agent whose job is to promote public health awareness in the community among other duties. However, Organisation X are doing it intensively and are able to visit each household of the beneficiary to see for themselves if people really put into practice whatever they are taught to do to improve on household hygiene. [Source: mjf-7]

In short, tables and maps that show frequency counts offer one important and useful but also selective and limited synthesis through which readers can gain insight into the data. They are also a device for opening up the data to further scrutiny and peer review: an alternative to the tendency for qualitative analysis to leave a chasm or 'black box' between summarising what data was collected (how many interviews etc.) and advancing arguments about what they revealed.

However, it is also important to emphasise that while frequency counts provide some indication of the weight that readers may give to evidence in modifying their prior views, they cannot be interpreted as an indicator of the strength or importance of a link across a wider population in any mechanical way. There are two reasons for this: first, they are not usually based on representative samples; and second, codes are ascribed to a range of similar but distinct statements that do not necessarily mean exactly the same thing (in terms of 'construct validity'). Hence while frequency counts can add usefully to weighing up the evidence generated by a QuIP, they need to be interpreted with care. For example, it is advisable to be careful about generalising from QuIP counts by using percentages (80% of respondents claimed x). Instead, it is better to rely on visualisations and descriptions of the number of sources who claimed particular causal connections, thereby not opening yourself up to unnecessary criticism for extrapolating from a dataset which is not representative or statistically significant.

Another limitation of causal mapping as a way to collect and weigh up evidence of change, even in the absence of intentional bias, is that the stories people tell can only be as rich as their willingness and ability to explain them in words, and the strength of social norms about what sort of stories it is normal to share and at what level of detail. Part of the art of the analyst at the reporting stage is to identify causal processes that appear contingent on additional (confounding) factors, which respondents may or may not mention. For example, farmers may attribute increased crop yields to new seeds without always mentioning that this was <u>also</u> made possible by timely and sufficient rainfall. A discussion about 'missing' elements in causal maps may therefore arise in sensemaking or triangulation workshops, and these can be incorporated into the final report. However, the maps themselves should only ever represent what was reported by sources in the dataset. In this way the process of deriving empirical causal maps from a given database remains transparent and open to peer review - even if it needs to be sense-checked alongside other data points. More on this in the next section.

4. How QuIP analysis relates to other forms of enquiry

So far, this note has described what kind of evidence a QuIP study can derive from narrative text data. We now consider how the QuIP relates to other approaches to producing evidence of causation. When does a QuIP offer an alternative to other approaches, or how can it complement or combine with them? We also clarify how the QuIP relates to the widely used distinction between qualitative, quantitative and mixed methods of research and evaluation.

Stand-alone QuIP studies. On its own, a QuIP study can generate useful evidence of how diverse respondents perceive changes taking place in their lives across a range of issues. To elaborate, the combination of causal maps, tables of citation counts and selected quotations a QuIP generates throws light on two kinds of complexity. First, there is the *substantive complexity* associated with systems that are fast changing, have many 'moving parts' or factors, and are hard to define (i.e. have open or porous boundaries). Second, a QuIP can be useful in revealing *cognitive complexity*, or variation in how different stakeholders perceive the system. The success of many public interventions (e.g. to encourage changes in consumption behaviour) depends not only on understanding what is going on in an absolute or objective sense, but also how others perceive the issues differently, and hence how best to engage with them. Substantive and cognitive complexity often go together, but a focus on one can lead to neglect of the other.

In practice, of course, no study exists in isolation. More specifically, a stand-alone QuIP study will also feed into organisations' ongoing processes of monitoring, performance measurement, performance management and learning. QuIP studies may be particularly useful to assist in **explaining variation** in performance indicators over time – e.g. why are some units/programmes doing so much better than other units? In the short-term they may help **inform specific programme or policy adjustments**. In the longer-term, they can contribute to **incremental adjustments** in the **underlying theories** and **shared mental models** underpinning not only what organisations do, but also their self-identity and purpose.

Having suggested that the evidence produced through a QuIP can have a quite profound influence on an organisation it is also important to highlight four of its limitations:

- Because it relies on open-ended narrative statements addressing issues with many moving parts it is not appropriate for generating estimates about the precise magnitude of specific causal links.
- The credibility of the causal maps generated by a QuIP depends on the credibility of its sources - they make no claim on their own to represent 'objective' truth, only respondents' perceptions of this.
- The scope for generalising from the evidence produced depends on source selection. While QuIPs often draw on more than the basic 24+4 dataset discussed here, they do tend to rely on relatively small sample sizes, and this limits the scope for credibly generalising (see Copestake, 2020, for a discussion of this).
- Because QuIP studies focus on outcomes, they do not systematically examine how a particular project or intervention was implemented.

Given these limitations, there is a lot of scope for using data generated by QuIP studies alongside evidence from other sources in complementary ways. Here we pick out six leading examples.[2]

[2] For a more comprehensive review of how QuIP relates to other approaches to impact evaluation see Chapter 2 of *Attributing Development Impact: the QuIP case book*, by James Copestake, Marlies Morsink and Fiona Remnant, 2019.

1. **Pilot studies**. QuIP-generated evidence can help to clarify concepts, select factors (variables) and prioritise the causal pathways to be investigated subsequently in greater depth or on a larger scale, including through use of surveys.[3]

2. **Theory-led process tracing.** A QuIP study can be one useful component of process tracing and contribution analysis that aims to identify packages of necessary and sufficient conditions for the achievement of specified outcomes. Such research entails coming up with a range of possible theoretical explanations or mechanisms for a specified outcome, followed by empirical tests to help decide which explanations are most likely to apply to different situations. QuIP studies can generate this kind of empirical evidence. Citation counts can also inform the process of Bayesian updating: raising or lowering confidence in prior explanations of what is happening.

3. **Mixed methods impact evaluation**. The classic design comprises open-ended interviews and focus groups (the qualitative 'small n' component) alongside a large-scale survey (the quantitative 'large n' component) which could be a randomized controlled trial, for example. The quantitative element aims to generate precise and valid estimates of the average or typical statistical association between key input variables or 'treatments' (X) and key outcomes (Y) across a defined population. If well designed, causality can also be inferred from these associations. The qualitative component, which could be one or more QuIP studies, helps to illuminate the possible causal mechanisms driving the observed changes, and contributes to understanding variation in the impact across the population.

4. **Process evaluation.** These typically combine a thorough review of documentation about a specific project with key-informant interviews to identify and explain the reason for progress (or lack of) in implementing a project as planned. They are conducted by one or a team of subject specialists, often over a relatively short period of time. Process evaluations often struggle to collect and analyse meaningful feedback from clients, end-users and intended beneficiaries. This gap can be filled by including a QuIP study as one component of the process evaluation.

5. **In-depth follow-up studies.** In addition to conducting a QuIP alongside other studies, an additional possibility is to utilise it as a way of following up on a particular question or issue generated by a previous study. This might arise, for example, where interpretation of a large-scale survey is proving difficult or contentious; or it might help to understand reasons for variation in the experience of different participants in a project. One advantage of this approach is that the information generated by an earlier baseline or repeat survey provides a strong foundation for purposeful selection of respondents for the QuIP. This fits well with the tradition of 'realist evaluation', the role of the QuIP being to assist in identifying 'context, mechanism, outcome' configurations experienced by different respondents.

6. **Participatory sensemaking**. The evidence generated by QuIP studies is generally primarily intended for sharing with managers and commissioners of projects, programmes and organisations, particularly where there are large gaps (geographical and cultural) along the financing chain. These can be linked to what some economists call 'information asymmetries between principals and agents'. However, the interpretation and use of QuIP evidence need not

[3] A special case of this is the use of QuIP to inform production of the causal maps needed in the design of quantitative impact assessments, as discussed by Judea Pearl and Dana Mackenzie in *The Book of Why* (2018). The kind of maps ('directed acyclic graphs') used by statisticians to model causal relationships differ from the kind of causal maps generated by QuIP studies in particular because the former model quantitative causal relationships between variables. Nevertheless, by collecting evidence of perceived causal mechanisms QuIP causal maps can contribute to the production of the causal models needed to inform statistical analysis.

feed only 'upwards' along the financing chain. Appropriately visualised there is a lot of scope for sharing them with other stakeholders too, including feeding back to those interviewed. An alternative to this is participatory causal mapping. This brings together stakeholders to agree on a map and can contribute to promoting collaboration and building a common understanding of a system or issue. QuIP, in contrast, permits more detailed analysis of how cognitive causal maps vary between stakeholders.

5. QuIP: qualitative or quantitative?

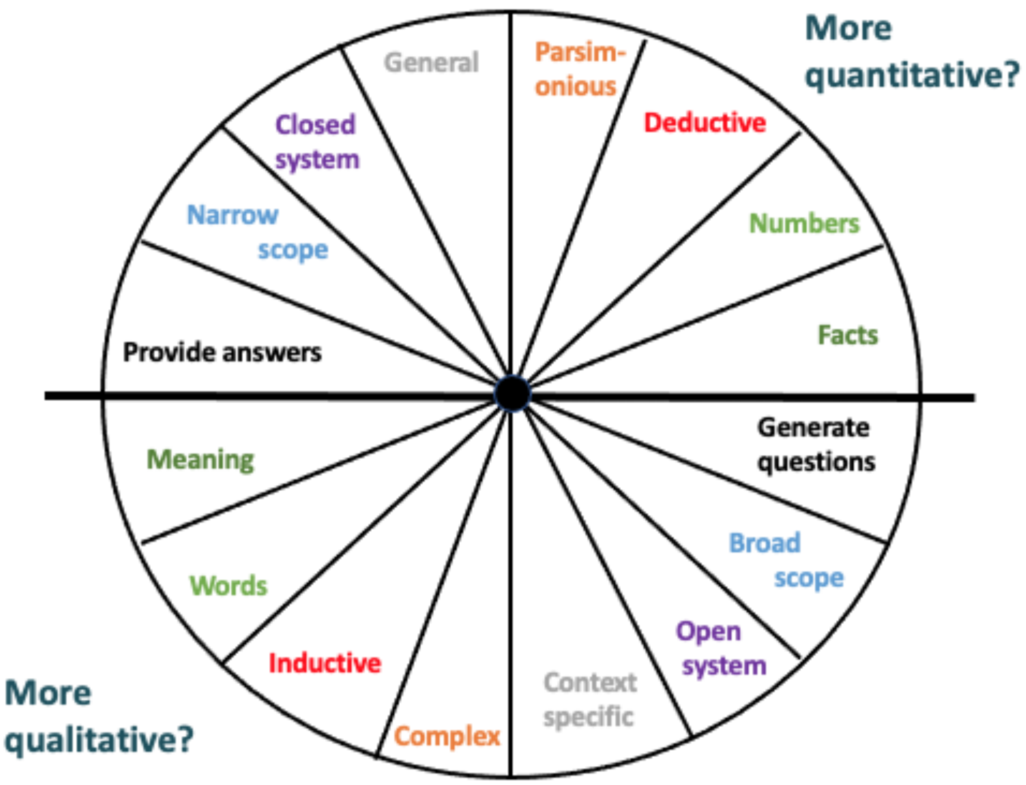
It is evident from its very name that the QuIP is primarily a qualitative approach to generating evidence of causation, but the issue of how QuIP relates to the distinction between qualitative and quantitative research is more complex than that, and depends on precisely how the two terms are defined. Here we explore just three different ways of doing so.

The first way of drawing the *qual/quant* distinction is to label specific research tools as one or the other. This is what we mean when we say the QuIP is a qualitative tool, and then go on to discuss how the QuIP can be incorporated into larger mixed method studies as we have above.

A second way to distinguish between them is as fundamentally different ways of thinking or research paradigms. Broadly, *quant* studies collect numbers and aim to generate 'objective' facts, whereas *qual* studies collect words and interpret them in pursuit of wider 'subjective' meaning. The QuIP was developed partly as a strategy to counter the dominance of a broadly quantitative mind-set in the field of impact evaluation, which at times seemed to belittle the power and importance of self-reported attribution of causal links over what could be observed and measured statistically. However, distinguishing between qualitative and quantitative this sharply can be awkward for those of us who collect numbers <u>and</u> words, and who often doubt what others claim to be objective facts while also recognising that it is possible to achieve a high level of 'intersubjective' consensus about what something means within a particular research community. Indeed, this is perhaps where most of us sit most of the time. More fundamentally, it is part of the mystery, miracle and power of our brains that they can both precisely select and codify complex information as 'facts' <u>and</u> generate feelings about them at the same time.[4] Figure 8 further warns against drawing a sharp distinction between quant and qual, because it can contribute to conflating and confusing a range of fundamentally different attributes of the process of establishing causal claims that do not need to go together.

[4] See https://www.ted.com/talks/iain_mcgilchrist_the_divided_brain

Figure 8: Deconstructing the conceptual distinction between 'qualitative' and 'quantitative' research.



There is a third and more profound way of negotiating the *qual/quant* field: one that gets inside individual research tools. An important step in quantitative research is to select and codify information about a complex world to facilitate statistical analysis – literally turning words into numbers. This is exactly what QuIP does through the process of coding causal claims, as illustrated by Figure 1, which also illustrates how we can also decode abstract data by 're'-presenting it in a qualitative synthesis, such as a summary causal map. In short, the quant/qual distinction reflects more granular research processes of selection and codification of data (*quant*), and of reframing and synthesis of data (*qual*).[5]

The QuIP can be defined as mainly qualitative in this regard: it mostly seeks open-ended textual data, partly (and where ethically acceptable and practically possible) by blindfolding interviewers and interviewees about narrower purposes to which the data will be directed. But at the same time, data collection using the QuIP also entails processes of framing, narrowing, selection and simplification of the respondents' world and how they view and experience it.

However, the purpose of this paper has been to elaborate more on data use rather than its collection. Having selected a sample of respondents, framed the conversations and 'captured' their content (mostly in words, but also with some numbers) what do we do next? A pure qualitative research answer would emphasise immersion in a body of data by an analyst, systematic (but unavoidably subjective) extraction of core meaning from it, and an attempt to distil this meaning in words (or

[5] This is not a new thought. For example, see Moris and Copestake (1993), who define the distinction as follows: "…the distinction between quantitative and qualitative enquiry hinges less on the source of information than on the point at which information is codified, or otherwise simplified. Early codification permits rigorous statistical analysis, but at the same time entails introducing restrictive assumptions which limit the range of possible findings." *J Moris and J Copestake (1993) Qualitative enquiry for rural development: a review. London: ITDG. Page 1.*

music, pictures, movement) in ways consistent with the totality of the data and to connect with selected audiences. This entails some selective reframing, but also **triangulation of data from multiple sources** in search of a synthesis that does justice to the complex reality from which the research started. The way we frame the *quant/qual* distinction itself influences attempts to standardise and share how data is de-codified and synthesised.

6. Conclusion

QuIP analysis entails processes of codification and counting that give it at least a partially quantitative flavour. It also aspires to a detailed level of procedural transparency and replicability that is more akin to quantitative than to qualitative data analysis. However, it is strictly interpretive in the sense that it aims not to deliver definitive facts about what a group of respondents think, but a systematic and transparent interpretation of this data. Hence it does not shy away from acknowledging that the subjective perspective and positionality of the analyst also matters.[6] It is then down to the user of QuIP generated evidence to assess how much to adjust their prior expectations of impact in the light of the additional evidence it generates in support of different causal claims.

Causal maps and frequency counts of different kinds of drivers of change, causal claims and outcomes are a useful way of presenting this evidence – the frequency of repetition of a claim (or lack of it) does credibly affect the weight of evidence offered, even though frequency counts are weak proxy indicators of the importance of different findings. The quantitative flavour of evidence a QuIP serves up should not divert attention from the often much richer qualitative insights on offer. Nor does it undermine the qualitative and interpretive philosophy underpinning the QuIP as an impact evaluation method.

[6] For more on this see J Copestake, G Davies and F Remnant (2019) 'Generating credible evidence of social impact using the Qualitative Impact Protocol (QuIP): the challenge of positionality in data coding and analysis' in *Myths, Methods, and Messiness: Insights for Qualitative Research Analysis*, edited by B Clift, J Gore, S Bekker, I Costas Batlle, K Chudzikowski and J Hatchard. An edited volume of proceedings of the 5th annual qualitative research symposium at the University of Bath, UK.