

Qualitative Impact Protocol (QuIP) Annotated Guidelines

This document is an annotated version of the full Guidelines available in the book, *Attributing Development Impact* (see www.bathsdr.org/resources for a full version). This shorter version is designed to be more accessible to non-native English-speaking MEL staff and is available in different languages.

The Qualitative Impact Protocol (QuIP) was developed at the University of Bath in the UK to address the challenge of assessing the impact of interventions in complex and/or rapidly changing contexts in a way that is credible, timely and cost-effective. The QuIP relies on evidence of the causal drivers of change obtained through in-depth interviews with selected respondents whose stories are likely to be relevant to the theory of change being tested. It is particularly designed to reduce potential response bias and to address the challenges associated with analysis and presentation of qualitative data of this kind.

These guidelines aim to provide a detailed introduction to anyone planning to undertake a QuIP study. However, we recommend looking at more resources and information on training which are available at www.bathsdr.org.

A. Overview

Individuals and agencies who commit to actions with social and development goals need evidence about whether they are indeed achieving what they intended. Such actions may be referred to in many ways: as grants, investments, interventions, projects or programmes. For convenience here we use 'project' to refer to any of these. In all cases the actors need evidence to help them decide whether to carry on, to expand or to change what they are doing. They also need to inform those with whom they work, including people intended to benefit from the actions and those helping to finance it.

In diverse, complex and rapidly changing situations it is not obvious how best to obtain such evidence, and this depends to some extent on why evidence is most needed. Is it primarily to demonstrate that past actions worked, to identify specific ways to improve on-going activities, or to reflect on an organisation's underlying mission and vision? Is it more important to quantify the magnitude of impact, or to explain why this varied from person-to-person or from place-to-place? How credible does the evidence have to be, and what level of expenditure can be justified? There are many ways of answering these questions. Their strengths and weaknesses vary according to context, and no one method or approach outperforms all the others under all conditions.

Quantitative ways to assess impact can be rigorous and precise but suffer from many problems: they can be expensive, slow, emphasise average effects and say relatively little about how change takes place. QuIP looks at impact from the perspective of people on the ground, and what they report as most important to them. It is good for understanding impact in context, including explanations for variation in impact and the contribution made by an intervention in complex and changing situations.

The QuIP's main purpose is to serve as a reality check on whether the social effects of a planned

activity or set of activities on intended beneficiaries is as expected, or whether it is having any unintended consequences. It can also provide insight on other factors that are affecting hoped-for changes (some perhaps unexpected) or highlight variations across a selected group in the changes experienced and the perception of the causal drivers of these changes. The QuIP is versatile: it can be used in a relatively narrow way to confirm whether a specific intervention is working as anticipated – e.g. as set out in a project’s theory of change, but it can also be used to explore what is driving change as part of an organisation’s broader commitment to reflecting on its priorities, strategies and activities. Generally, however, it is not so useful for capturing the magnitude of changes; for this reason, some people refer to it as a way of assessing impact contribution rather than attribution. However, the QuIP can usefully assist in estimating the magnitude of possible impacts when used in combination with other methods.

In contrast to quantitative methods, the QuIP sets out to generate case-by-case evidence of impact based on narrative causal statements directly from intended project beneficiaries, without the need to interview a control group. Evidence of attribution is sought through respondents’ own accounts of causal mechanisms. This contrasts with methods that rely on statistical inference based on variable exposure to an intervention.

There are strong ethical grounds for asking people directly about the effect of actions intended to benefit them but doing so involves finding credible ways to address potential response biases. The QuIP does this by arranging for qualitative data collection to take place with as little reference as possible to the specific activity being evaluated, and by giving equal weight to all possible drivers of change in possible domains of impact. This is achieved by working, with researchers who are completely independent of the organisation responsible for the actions being evaluated. Indeed, where possible, researchers are ‘blindfolded’ at least from the details of project implementation and the theory of change behind its actions – and sometimes the identity of the organisation being evaluated. Evidence collected from respondents takes the form of ‘stories’ about causal drivers of change in selected areas of their life. Another researcher, the analyst (who will have access to the background information), then analyses these statements using a standardised approach to coding causal connections in the narrative, highlighting whether the reasons given for change explicitly or implicitly confirm or undermine the causal theory underpinning the intervention (or are completely incidental to it). Where possible, this analysis can then be compared with observed changes and monitoring data on project activities, helping to build a more detailed picture of what has really changed and why.

Another potential limitation of self-reported evidence of impact is that it is restricted to what respondents actually know and what they regard as most important. Use of the QuIP does not depend on believing that respondents are all-knowing, but that their experiences and opinions are insightful and important. It can be used alongside other forms of evidence to identify important cognitive gaps between different actors. For example, if project staff and intended beneficiaries do have widely contrasting perceptions then it is likely to be useful to know this.

When planning a QuIP study there are several different elements that can be altered in the methodology to meet the specific requirements of the project being assessed. Questions to consider include, why is a QuIP study being considered, by whom, and how they will use the evidence it generates alongside information from other sources? This will influence decisions about what other data might be needed, how the timing and sampling strategies will overlap, whether blindfolding is desirable or achievable, and who will be involved in each stage. The next section considers these questions and how to plan for a QuIP study.

B. Planning and designing a QuIP study

B1. To QuIP or not to QuIP?

The QuIP offers one solution to the attribution challenge; but it isn't appropriate in all situations and is often best combined with other methods to generate all the evidence that may be expected of an evaluation. It is important to manage the expectations of all involved about both its potential to add value and also its limitations.

The QuIP **does**:

- Generate insights into intended beneficiaries' *perceptions* of change and their understanding of why these changes have happened.
- Highlight variation in change within a population of intended beneficiaries and the reasons for these.
- Assist in confirming or refuting the theory (of change) behind a project in relation to specific intended beneficiary groups and areas sampled.
- Generate such data in a more credible way by reducing the risk of pro-project bias, through incorporation of an appropriate level of blinding (always adjusted depending on the research context).
- Use a qualitative questionnaire developed with the commissioner to explore perceived changes across a variety of livelihood and wellbeing domains.
- Employ experienced and skilled local researchers who conduct interviews with intended beneficiaries in an appropriate local language.
- Code and analyse interview data in a transparent, systematic and rigorous way using flexible thematic coding (for identifying different drivers of change and outcomes and the degree to which these can be attributed to the project).
- Enable and encourage users to refer back to source text data, by providing an annotated annex of all coded interview data and/or access to this digitally through a dashboard.
- Generate data that can be used in a wide range of stakeholder and 'sense-making' meetings, including with project staff and intended beneficiaries.

The QuIP **does not**:

- Provide results that are statistically representative of all intended beneficiaries. QuIP studies are designed to gain a deeper insight into changes occurring in purposively selected groups, and to permit cautious generalisation across the wider population.
- Guarantee to answer very specific questions about the impact of certain project activities. If the activity is considered important by respondents in a wellbeing domain covered in the interview (and not simply taken for granted) then the QuIP should pick up unprompted references to these project-related drivers. However, if project activities are relatively marginal to respondents' lives then a more direct and targeted line of questioning is required. However, gaining a better understanding of the broader context of change (including factors that contribute to or mitigate the success or failure of the project) may

still be useful.

- Measure the magnitude of impacts or provide detailed quantitative data. The QuIP focuses on the nature of impact rather than its magnitude. Some quantification of drivers of change and outcomes can be generated to summarise and visualise patterns and themes across the sample, but the data is not statistically representative. It may be useful to inform modelling that can simulate the magnitude of change, but other data will be needed with which to calibrate such models.
- Score or weight the overall success or failure of a project. Whilst the visualisation of coded qualitative data can make the evidence easier to digest and highlight patterns and outliers, commissioners need to be prepared to engage with the data, and where possible triangulate with evidence from other sources to make an overall assessment of the project and draw out recommendations for future action.
- Directly promote a more participatory approach to development, although findings can be used to promote reflection and learning among intended beneficiaries, and some respondents have also reported finding the interviews and focus groups to be useful and/or enjoyable opportunities for self-reflection.

B2. Who will be involved in carrying out the study?

The **commissioner** is the primary consumer of evidence to be collected, and responsibility rests with them to decide what sort of evidence they want, as well as when, where, how and why to collect it. Their main responsibilities include confirming the scope of the study, agreeing on the sampling strategy, providing relevant project documentation to enable sample selection, overseeing and supporting appropriate dissemination and use of findings, including ensuring the interpretation of QuIP data is integrated with evidence generated in other ways. Aside from this there are three main roles in a QuIP study:

- The **lead evaluator** is responsible for working with the commissioner, designing and managing the study, commissioning data collection from a field research team, and overseeing analysis and reporting. Contracting someone from outside the organisation to perform the role is likely to strengthen the credibility of the evidence produced but they may be an employee of the same organisation that is implementing the project if they are not directly involved in management of the project. Their main responsibilities include designing the questionnaires and sampling strategy, recruiting, training and managing researchers, overseeing data analysis (if not doing themselves) and producing a synthesis report to be used with key stakeholders. The lead evaluator will need to be familiar with the principles of qualitative data analysis and must also be in a position to manage the sub-contracting of the researchers who will conduct interviews.
- The **lead researcher** plays a key role in the QuIP process and is responsible for managing all aspects of data collection. They will typically be experienced qualitative researchers from the country where the evaluation is taking place, with a track record of conducting high quality fieldwork and recruiting, training and managing a team. This role can also be conducted by the lead evaluator if the circumstances are right; they must be able to maintain a commitment to the goal of enabling the authentic voices of intended beneficiaries to be heard. The main responsibilities of the lead field researcher include recruiting and managing an experienced team, taking responsibility for gaining access to the pre-selected sample of respondents, ensuring interview data is of a high standard,

maintaining good communication with the lead evaluator.

- The **analyst** is responsible for coding all the interviews using the QuIP approach to thematic qualitative analysis in appropriate software, analysing the coded data and pulling out key findings in preparation for report writing or de-briefing with the lead evaluator. The role of analyst and lead evaluator can be combined, but since they require very different skills and analysis can be very time consuming there is a good case for separating them where this skill set and availability cannot be combined - as long as they can communicate and collaborate effectively. An effective analyst must be able to immerse themselves in the data and identify and untangle often complicated causal claims and stories of change, both positive and negative. The analyst is expected to pull out the main findings from the data, construct the relevant tables and data visualisations and present these as the foundations for a QuIP report.

B3. When to carry out a QuIP?

Deciding when to schedule a QuIP depends on its relationship to the project being assessed.

- Early in the design phase, as a diagnostic tool for identifying drivers of change or testing the theory behind a proposed project.
- Early on or mid-way through a project, as a 'deep-dive' or 'reality check' to find out what intended beneficiaries think is happening, with time for course correction based on information gleaned – e.g. from individuals that monitoring data suggest are positive and/or negative deviants.
- After, or at the end of a project, to inform reflection on what worked and why (including the relevance, sufficiency and reliability of assumptions and theory underpinning the project), even when there isn't a baseline or control group to aid impact evaluation through statistical comparisons.

B4. How to select a sample?

There is no universal best practice method for selection of cases for a QuIP study since it depends upon many contextual factors. The most important of these are (a) the main purpose of the study, including its role in assessing an explicit theory of change, (b) availability of relevant data about variation in the characteristics of expected gainers and losers from the project, (c) availability of relevant data about variation in their exposure to project activities, (d) time and resource constraints, (e) how much data one analyst can manage. This section briefly explores these factors, and then outlines the sampling decisions needed prior to starting data collection.

(a) Main purpose of the study

Deciding who to interview, how many people to interview, and how best to select them requires clarity about what information is being sought, by whom and why. Neglecting this not only leads to poor practice, but also to misunderstanding about the quality of a study. For example, sample bias is not an issue for a QuIP study that deliberately set out to identify drivers of successful outcomes by interviewing positive deviants. Deliberately selective or explicitly biased sampling is, in this instance, fit for purpose.

More generally, differences in sampling strategy arise from whether the priority is to confirm and quantify the overall impact of a completed project on a defined population in relation to a predetermined set of measurable indicators and theory of change, or to explore what is happening in a more open-ended way – to improve implementation of an on-going project, for example. The QuIP is a relatively flexible and open-ended approach. Its primary purpose is to gather evidence of causal processes at play, not to quantify them. Deciding on the number of interviews and focus groups to conduct depends less on reducing sample bias than on assessing at what point the extra insight into causal processes gained from more data is unlikely to justify the extra cost. As a benchmark, a standard ‘single’ QuIP consists of 24 individual household interviews and four focus group discussions. But this may need adjusting for many reasons, including the time required to locate respondents. It is common to do a ‘double QuIP’ to draw sub-samples from two contrasting segments of the population, or even four groups to total 96 interviews.

(b) Contextual variation

Random selection of respondents across the entire population affected by the project is a good starting point for thinking about sampling for a QuIP study, but there are also good reasons for departing from it. For example, if there are good grounds for expecting impact to vary for different sub-groups, and we already have data that enables us to identify those sub-groups then there are good grounds for stratification of the sample. A project may cover two areas with marked geographical differences, justifying including a minimum quota of people living in each (e.g. urban and rural areas, irrigated and non-irrigated villages). Stratification of the sample on these grounds is an art that depends on prior thinking about what contextual factors are most likely to be a source of variation in project outcomes. Where baseline and endline monitoring data has already been collected and analysed then there are additional possibilities for QuIP sample selection. For example, quota samples can be selected for ‘positive deviant’ households that have experienced rapid improvement in key indicators to find out more about the drivers of their success. Conversely there is a case for deliberately biasing the sample towards households that have done badly, to learn why. A third option is to do both to be more confident about picking up the full diversity of causal changes experienced by households. Or a double QuIP might quota sample four groups: richer and improving; richer but declining; poorer but improving; poorer and getting worse. In all cases the number of interviews that it is worth conducting depends not only on minimising sampling error, but also on the marginal benefit (in terms of extra evidence of key drivers of change) obtained from each extra interview.

(c) Exposure or ‘treatment’ variation

This refers to variation in how project activities are expected to affect different people, including those who receive different packages of goods and services. In addition, there are those who may only be affected indirectly: because their neighbours are affected and may share things with them, for example. If data is available on variation in who directly received what and when, and it is expected that these differences will have different causal effects, then there is a case for stratifying the sample to ensure it reflects a range of treatment exposure. Impact assessment using the QuIP does not require a control group of people completely unaffected by the project. There may nevertheless be an argument for interviewing some people unaffected by the project (but similar to those affected by it) in order to explore whether they volunteer different or additional drivers of change.

(d) Time and resource constraints

A third reason for departing from pure randomisation in sample selection is to cluster respondents geographically to reduce the time and cost of data collection. One way to do this is to adopt two stage random sampling, with the first stage based on geographical units (e.g. villages, districts or census areas) listed according to some known criterion that is likely to be an important source of variation in project outcomes (e.g. distance from a main road or market centre; agro- ecological zones). One locality is then selected at random, and additional localities are selected by counting X down the list, where X is the number of localities divided by the desired sample number. For example, if there are 40 villages with an equal number of intended beneficiaries in each, and it is agreed to sample four of them, then every 10th village should be selected from a random starting point on the list. In the second stage the procedure is repeated, except starting with a list of all intended beneficiary households in each selected village.

Ultimately, budget constraints (dictated by factors beyond the control of the lead researcher or even the commissioner) may also limit the total number of interviews and focus groups that a QuIP study can cover. The challenge is then to make decisions that maximise potential value, subject to this constraint. This is less precise but no less reasonable than using power calculations to work out the minimum 'required' sample size for estimating the value of a key indicator to an acceptable level of statistical significance.

(e) Absorptive capacity of the analyst

An additional influence on sample size and selection is the limit to how much data the analyst can immerse themselves in, yet still code comprehensively, systematically and inclusively. Going beyond a double QuIP is likely to stretch all but the most gifted and experienced analyst. If a larger sample is nevertheless justified then parallel or consecutive QuIPs can be conducted and analysed separately, and the reports can then be subjected to synthesis or meta-analysis.

Before deciding on an approach to case selection, it may be useful to consider these questions as a team:

- Is it more important to assess the *typical* experience of intended beneficiaries, or to focus on the diverse experience of more narrowly socio-economic groups, or those exposed to different 'treatments', or who appear from monitoring data to be doing particularly better or worse than others?
- Is overlap with samples used for other studies useful? Or is it important to avoid intended beneficiaries who have already been interviewed under other studies to avoid survey fatigue?
- Is it useful to collect information from individuals or groups who were not intended beneficiaries (e.g. those who may benefit or be adversely affected indirectly)?

B5. To what extent will the researchers need to be blindfolded?

Blindfolding – including double blindfolding – can help to reduce the risk of pro-project bias and hence enhance the credibility of findings. But the extent to which the field researchers are blindfolded will depend on your aims and the context of the study.

- Double blindfolding is only possible through involving a third party, in order that the field research team can be recruited, trained and supported without knowing the identity of the organisation implementing the project or commissioning the study.
- Partial blindfolding may be more appropriate – e.g. a trusted team of researchers might be recruited directly by a commissioner, but without being given information about the project being assessed.
- By not blindfolding them a trusted team of researchers may be able to obtain more detailed and relevant information about the project; their professional expertise and integrity may be more than sufficient to ensure they are impartial and do not prompt respondents to respond to questions in line with prior understanding and interests.
- No blindfolding may be necessary if it is impractical, unethical or dangerous to blindfold either interviewers or respondents. It is still possible to focus instead on designing an open-ended and exploratory questionnaire, positioning the study in a broader context, and encouraging respondents to refer to this broader context when thinking about drivers of change.

Any research involving people as participants or respondents must be based on ethical principles. Blindfolding respondents raises particular ethical questions that need to be carefully assessed prior to each study. Blindfolding does not have to be complete and should not be permanent; this should be temporary blindfolding as an appropriate means to a beneficial end. Organisations commissioning QuIP studies are encouraged to include triangulation, feedback and ‘unblindfolding’ workshops to which field staff and respondents can be invited once the data has been collected and analysed. Decisions about precisely how much detail will be hidden and how much revealed can be decided at the design stage, along with agreement on ethical principles and procedures concerning confidentiality and anonymity.

B6. What form do the interviews take?

The QuIP employs two main data collection instruments: semi-structured household level interviews and facilitated focus group interviews. The questionnaire for both are based on a series of livelihood and wellbeing domains designed to cover outcomes specified in a project specific theory of change. For example, a project designed to promote household-farm livelihoods, food security and nutrition might include domains for:

- Food production
- Food consumption
- Income
- Cash spending
- Intra-household relationships
- Inter-household relationships
- Overall wellbeing

Questions are designed to stimulate discussion in an open way, with lists of supplementary questions available to sustain and deepen conversations about changes observed by the respondent and the reasons behind them. Closed questions follow open-ended discussion of a domain and are a useful way of drawing discussion of it to an end before moving onto the next.

Example questions for a domain on food production

Open question

- Please tell me if and how your household produces food has changed in the past two years.
 - *Do you do more or less of anything?*
 - *Are you doing anything differently?*
 - *Have you taken up any new activities to help you produce more food?*
 - *In which seasons have changes been most pronounced?*
 - *What are the reasons for these changes?*

Closed question

- Overall, how has the ability of your household to produce enough food to meet its needs changed in this time? [Improved, No change, Worse, Not sure]

Interviews typically take around 1-1.5 hours, so sufficient time needs to be allowed for research teams to find respondents, introduce themselves, conduct interviews and then preferably write notes up on the same day (from recordings where possible, or use the services of a second notetaker present at interviews).

B7. Data analysis and presentation

A common issue with qualitative research and impact assessment is how to organise and make sense of large quantities of textual data, and to do so in a way that is transparent, so that generalisations drawn from it can be peer reviewed. These were the driving features behind the approach to data analysis developed for QuIP data. The process can be divided into five steps:

1. Familiarisation with all the data by reading and rereading it
2. Allocation of segments of the texts to different codes
3. Identification of wider themes, stories or arguments that may combine different coded elements together
4. Backchecking these themes, and the clusters of coded data supporting them, against the original data and with the commissioning team
5. Reporting findings to others in a credible and straightforward way, without losing sight of the richness of the underlying data

However, this process is rarely strictly linear, step 5 serving as a particular and important reminder that the analytical process is iterative. At the same time, the QuIP does also involve more tightly structured tasks, thereby distinguishing it from even more fluid ways of doing thematic analysis in social research.

One of the more mechanical steps is to analyse the closed questions about each domain. An overview of these results is easy to produce automatically, as illustrated below. This enables both the analyst and users of the study to gain a quick sense of who the interviewees were and what their perception of change was, within a specified period, across all domains. However, even this data can be presented and interpreted in many ways. For example, patterns can be revealed by ordering the list according to different socio-economic characteristics (e.g. by age, gender, location and/or wealth group). The data can also be triangulated against changes measured using quantitative baseline and endline monitoring data.

This initial analysis provides a useful profile of the sample, and experience of change across it, but does not reveal anything about the causal processes behind observed changes. To get at this, the QuIP analysis entails coding segments of the narrative data that make causal claims (e.g. 'X caused Y', or 'Y happened because of X and Z').

QuIP text analysis is based on two well-established social science approaches: Qualitative Data Analysis and Causal Mapping; coding and summarising the data either 'deductively' using predetermined themes, or more 'inductively' by identifying repetitions and patterns - and using this coding to build up causal maps. QuIP coding involves more tightly structured tasks, thereby distinguishing it from more fluid ways of doing thematic analysis in social research. We recommend coding only segments of the data that make causal claims (e.g. 'X caused Y', or 'Y happened because of X and Z'), and using your coding to flag:

- **drivers of change / influences** - based on inductive classification of the reasons behind any change or outcome;
- **outcomes / consequences** - also based on inductive classification, and allowing for limitless linked driver-to-outcome sets to be classified; one driver leading to an outcome, which in turn drives another outcome, or one driver leading to multiple outcomes simultaneously
- an **attribution** claim - deductive coding based on your theory of change (unless your study is purely exploratory); to what extent does the driver of change in the story implicitly corroborate or challenge the theory of change, or is it incidental but potentially significant?

This approach to coding enables analysis of respondents' reported experiences and how different drivers may have interacted to mitigate, or help intended change. Unlike the field researchers, QuIP data analysts need to be fully briefed about details of the project in order to code for attribution. Their task is to assess how the data relates to the project's theory of change according to whether the respondents (a) **explicitly** attribute impact to project activities, (b) make statements that are **implicitly** consistent with the project's theory of change, (c) refer to drivers of change that are **incidental** to project activities. Statements can also be coded according to whether respondents described their effects as positive or negative.

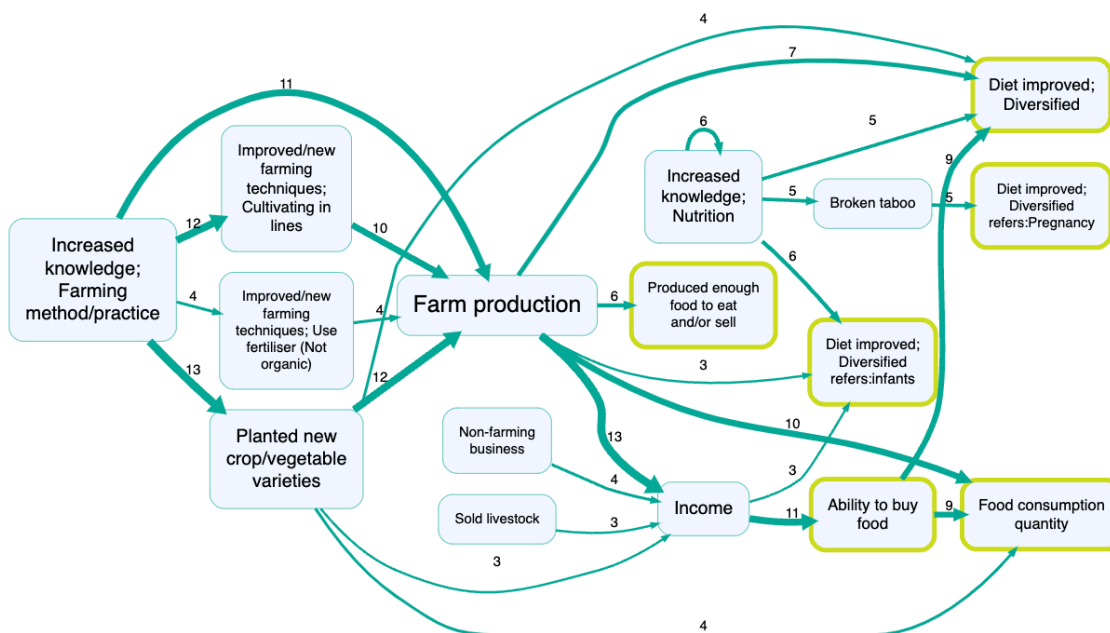
Various qualitative analysis software packages are available on the market, and even Excel can sometimes do what you need, but the QuIP's emphasis on causal connections led Bath SDR to invest in the creation of software (www.causalmap.app) which makes it easy to code cause and effect in stories of change, with maps generated automatically as you code.

Analysts highlight quotes within the narratives, and for each quote, identify a pair of causal factors: the cause and the effect. As the analyst continues to identify causal claims within the narratives, they will re-use existing factors mentioned by the same respondent or other respondents. In this way, a causal map can be built up showing all the links between all the causal factors. The result can be a rich network or map with many hundreds of causal factors and causal links, summarising the stories told by all the respondents. This map can then be aggregated and filtered to show particular aspects of the stories, and query how stories from respondents with different characteristics may differ.

Visualisations and calculations in the Causal Map app can help to answer:

1. Is there evidence that the programme 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 and frequency counts as seen by different age groups, gender etc.?

Example causal map looking at all drivers leading to changes in diet



Number of sources marked over the links. Factors outlined in green indicate the targets of the filter.

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 respondent’s original words. One aim of QuIP reporting is to encourage the reader to get involved with respondents’ original statements and read them in context.

Data from causal maps can be queried and analysed to very powerful effect and used to generate other tables and visualisations.

B8. Use of QuIP data and analysis

The data analysis described above can be adapted and taken further in numerous ways. The summary tables and maps are typically incorporated into a written report that also pulls out quotes from the source narrative data to illustrate and elaborate on key findings. However, use of findings does not have to rely on written outputs. For example, if trained staff from within the commissioning organisation do the coding themselves then internal learning starts even before the analysis is complete. Interactive dashboards can also be used to structure feedback meetings with project staff, individual respondents and other stakeholders.

QuIP analysis takes its cues about which outcomes are important from the respondents themselves, so it is logical to involve them in triangulation workshops, enabling those attending to challenge, corroborate and complement findings. This both serves a quality assurance function and deepens understanding of what changes took place for whom, how and why.

Key interpretive questions include:

- To what extent are findings consistent with both transmission mechanisms and intended outcomes set out in the theory of change?
- What evidence of processes and outcomes is generated that is not consistent with the original theory of change, and how can these be explained?
- What scope is there for generalizing reasonably from findings to the whole project, taking into account characteristics of the whole sample of intended beneficiaries and of the sample of those interviewed?
- What explains differences in intended and observed processes and outcomes of the project and what are the implications for future activities?
- Is the data consistent or at odds with quantitative monitoring data, as well as data collected from other sources (including meetings with project staff)? How can differences and similarities best be interpreted?

The commissioner should be provided with a report which will include core summary tables and other data visualisations picking up on the most interesting patterns in the data, appended by coded transcripts (and the data dashboard where relevant) which make it easy to find the source data. This ensures that all the data is available rather than only the quoted extracts selected by the evaluator, and that there is a clear reason for any selected extracts.

Once analysis has taken place (and if the research team are unlikely to be asked to participate in further blindfolded studies) then a powerful final stage of any QuIP study is to organise one or more fully unblindfolded triangulation or sensemaking workshops involving project staff, the

research team, respondents and other relevant stakeholders. This ensures greater transparency and enables researchers to offer their own interpretation of the findings, drawing on what they wrote down, their direct field observations and wider experience. The discussions from such workshops can be useful for putting the QuIP findings into a broader context and starting to draw up internal recommendations for practical action.

Negative or unexpected findings may be a source of internal tension, with some staff or stakeholders preferring to see them buried or dismissed without proper reflection (an issue that can also emerge in discussion of draft reports). Such tensions can be viewed as obstacles to completion of studies and make unplanned and unwarranted demands on time and resources, but they can also in themselves be powerful learning opportunities.

An alternative follow-up initiative is for commissioners to report anonymised findings back to the respondents who were interviewed for the study through one or more focus groups or sensemaking workshops. This provides an opportunity to thank respondents for their participation, and to explore how they interpret the findings in more detail. Uncertain findings, and specific questions which were not answered in the original interviews can be explored further, and scope for follow-up project activities discussed.