QuIP and the Yin/Yang of *Quant* and *Qual:*

How to navigate QuIP visualisations

1. Introduction

When is a table, chart or diagram qualitative and when is it quantitative? Can they be both at the same time? Does it matter what we call them?! The QuIP is self-defined as a qualitative approach to impact evaluation and research. But it's not quite as simple as that; defining and negotiating the *qual/quant* interface rarely is. It's an issue that can be fiercely dull, but also one that can't be entirely avoided - not least as a riposte to others who use the distinction in different, and often unhelpfully simplistic ways to support their views about what does and does not constitute credible evidence. Mostly, the *quant/qual* issue arises when discussing data collection, but it can also arise in discussion of data presentation. This paper first contrasts three very different ways in which the distinction is used: to distinguish between big paradigms, specific tools and more granular research processes. Drawing particularly on the third of these it then reviews how quantitative (codifying) and qualitative (de-codifying) processes are utilised in analysis and presentation of data using the Qualitative Impact Protocol, or QuIP. The paper also includes a glossary of key terms used in QuIP as an appendix.

2. Defining the *qual/quant* interface

One way of drawing the distinction between quantitative and qualitative is as mind-sets or research paradigms. Broadly, *quant* studies collect numbers and generate facts, whereas *qual* studies collect words and pursue meaning. This distinction can be awkward for those of us who collect numbers and words, and who often doubt what others claim to be facts while also recognising that it is possible to achieve a high level of consensus about the meaning of at least some evidence. 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' and generate feelings about them at the same time.[1]

A second 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 mixed methods: or its use to complement and triangulate both *quant* tools and other *qual* tools. QuIP can be nested within larger studies. It can serve an exploratory function that is best followed up by collecting more narrowly specified and precisely coded data across a larger sample, opening up scope for more complicated statistical analysis. Or QuIP can take its cue from prior analysis of quantitative data, and be used to dive deeper into how people explain and evaluate the reasons for the correlations, changes and patterns that it reveals.

But there is a third and perhaps lesser used way of negotiating the *qual/quant* field: one that gets inside individual research tools, and this is what is emphasised here. We've already suggested that an important step in quantitative research is to select and codify information about a complex world in order to facilitate statistical analysis. This can be very powerful, but always begs more qualitative questions about whether we've framed the problem in the best way, for whom, and for what purpose. Have we framed a question in a way that is unhelpfully simplistic or even biased by our disciplinary perspective, values, interests, goals and entry points? In other words, the quant/qual

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

distinction reflects more granular processes in research of selection and codification of data (*quant*), and of reframing and synthesis of data (*qual*).[2]

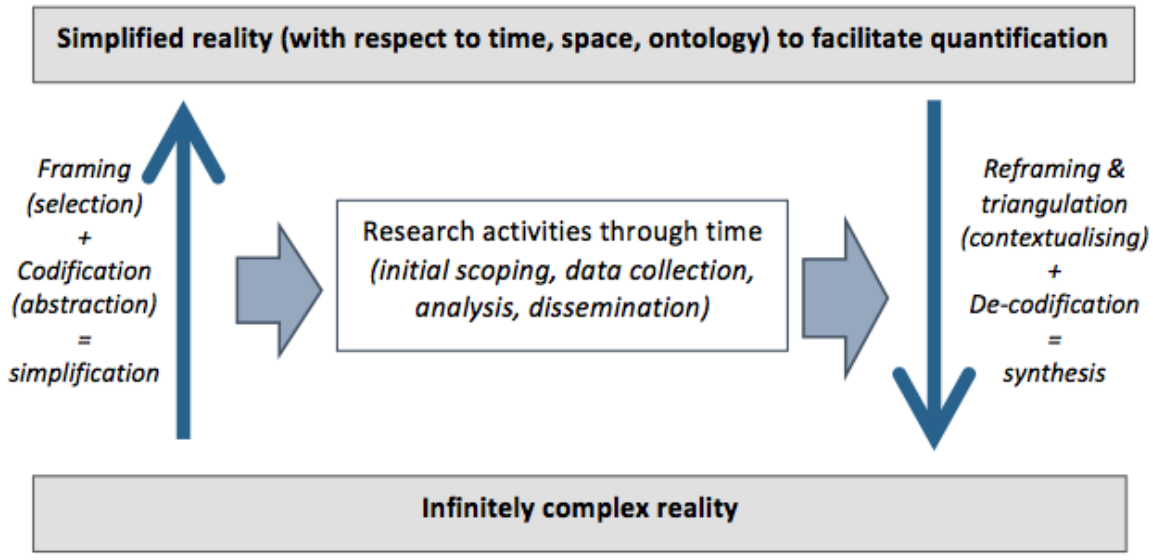
The QuIP can be defined as mainly qualitative in this regard: it mostly seeks open-ended textual data, partly (and where ethically acceptable and practically possible) by blindfolding interviewers and interviewees about narrower purposes to which the data will be directed. But at the same time, data 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 is 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 music, pictures, movement) in ways that 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.

This may sound more like a mediation on how to be an artist than a social scientist, but perhaps that's because the obvious differences between artists and evaluators in framing a task (and in bolstering claims to truth through procedural transparency and peer review) should not hide strong aesthetic commonalities between the two roles. And the *quant/qual* issue intrudes again precisely over attempts to standardise and share how data is de-codified and synthesised.

The diagram below attempts to summarise the framing and reframing of data in a QuIP, demonstrating how complex reality from open-ended questionnaires is simplified initially in both the framing and coding processes, but once in a more accessible form can then be re-framed through triangulation with other data, and de-codified by giving easy access to the full narrative data behind the numbers.

Figure 1



[2] 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.*

3. Visualisation and analysis of QuIP data

This section enters into the detail of how QuIP data is coded, analysed and presented. In brief, we start with an overview of how respondents perceive change in their lives, across selected domains. We then summarise the extent to which these changes can be attributed to a specified 'project'. Third, we set out in more detail the drivers of these changes. Fourth, we may go into more detail about the perceived change in outcomes and the causal pathways behind them.

The core of a standard QuIP comprises text data from interviews with 24 individuals and four focus groups – let's refer to these as 28 respondents. This is loosely organised by a variable number of predetermined outcome domains – let's say for simplicity there are ten of these. That means that we have a problem of how to capture and convey what is most useful about the field of activity the commissioner is interested in across 280 sets of discrete pieces of textual data. In each set we are looking particularly for causal claims that link the outcome domain to one or more drivers of change. We may also distinguish between more than one outcome (and indeed level of outcome) in each domain.

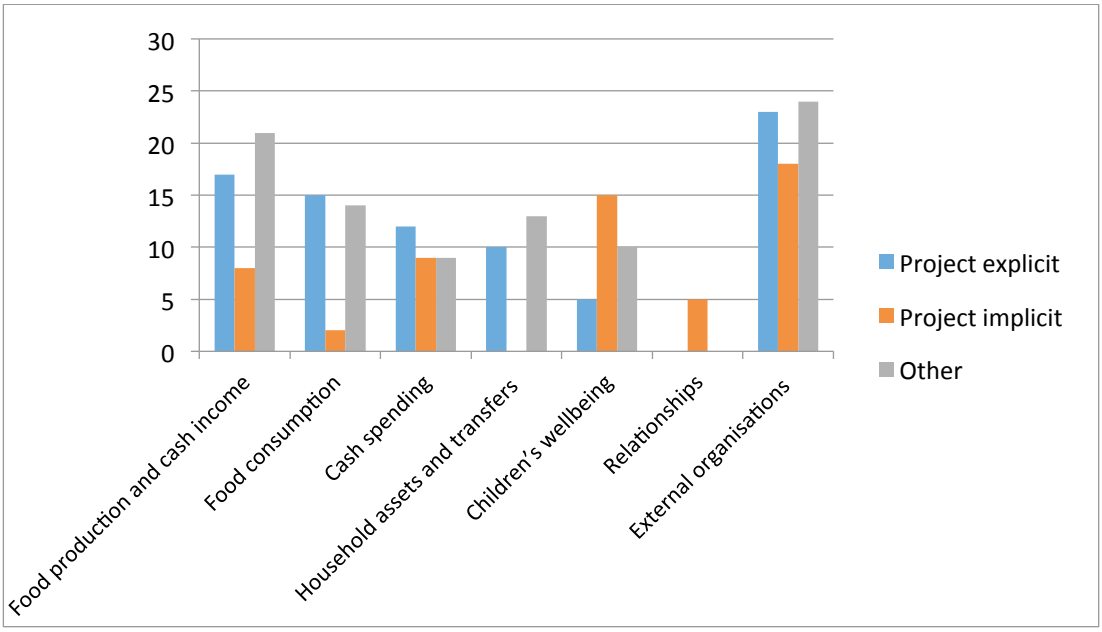
For each domain we also collect a small number of closed questions to allow each of the 24 individual respondents (but not focus group participants) to indicate whether change in the selected domain over a carefully specified period has been positive or negative for them. Reporting on findings starts with this, using a single table: one row for each person and one column for each question. Although the numbers only reflect respondents' broad perceptions, this is the closest we get to narrowly pre-coded quantitative data. However, the table is constructed partly to permit rapid synthesis. Are the cells mostly positive or negative? Are there patterns in the responses – e.g. according to respondents' age, gender or place of residence? Here we are clearly switching from coding (a *quant* task) to synthesis based on visualisation (a *qual* task). The intention with this table is to gain an overview of respondents' experience of change in selected domains, but to do so in a way that highlights variation in experience and doesn't hide individual perception behind aggregate statistics. An extract from a closed question table is below.

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	=	=	+	+	=	-	+	+	+	-
UFFC-3	Middle	-	-	-	-	-	-	-	+	+	+

If this data is an appetiser, then the main dish in a QuIP report is the presentation of the 280 sets of open-ended narrative data; and the major challenge is to avoid giving consumers indigestion! The first analytical step (coding) is to identify, classify and count different sorts of causal claims (linking outcomes back to what respondents perceive to be their main causal drivers). Given that the QuIP is designed as an evaluative tool, or reality check on whether the theory of change behind a project is realistic or not, then a logical place to start is with attribution. How many casual claims have we identified that link each outcome area back to causes or drivers of change that explicitly refer to the project, are implicitly consistent with the project's theory of change, or completely incidental to it? Again a table that counts and colour-codes explicit, implicit and incidental drivers of positive and negative change by domain provides a visual synthesis of where the bulk of evidence about the projects impact (or lack of it) resides. This data can also be presented using bar charts or histograms.

Figure 3: Example - Attribution of positive change across domains



Of course, this form of presentation does violence to the richness of the textual data and should be 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 (having hopefully first immersed themselves by reading across all 280 sets of data) regards them as particularly insightful or interesting in their own right. In short, tables that show frequency counts offer one important and useful but also selective and limited synthesis through which readers can gain an insight of what is in 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.

To this end the data presented in a QuIP report will contain not only numerical summaries in the form of charts or other visualisations, but constant reference back to the source data in the form of both quotations and tables which highlight where households fall across a range. These tables give each number an identity, enabling some eyeballing of different respondent types and also allowing easy reference back to the source data which is sorted by attribution code and by domain.

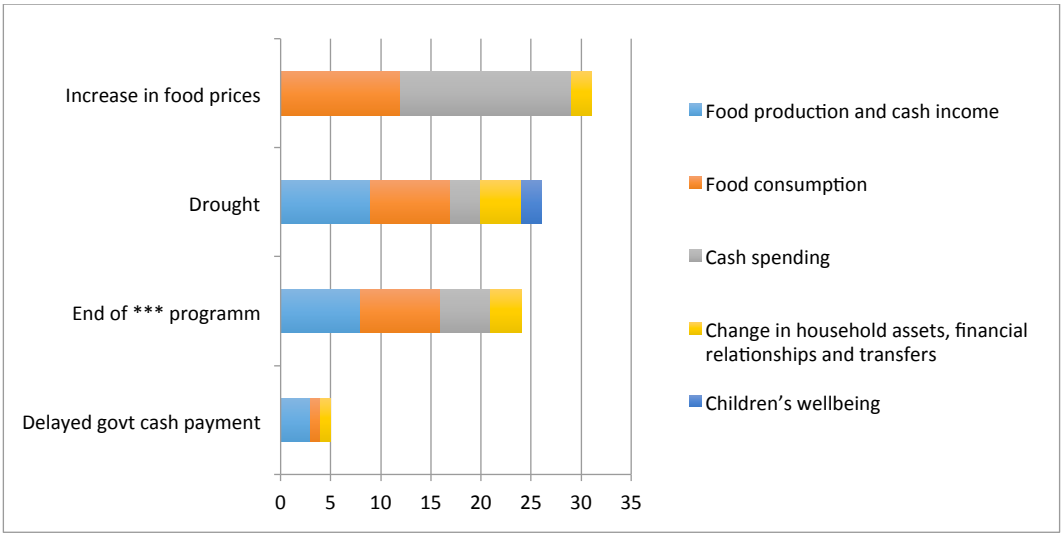
Figure 4: Example Attribution frequency of positive outcomes by domain with respondent codes

	Explicit attribution of positive change to ***	Implicit attribution of positive change to ***	Positive change attributed to other source
Food production and cash income	DHMC-1 DHFC-2 DHMC-3 DHMC-4 DHMC-6 DHMC-11 DHMP-12 DHFP-9 DHFP-10 DHMG-4 UEMC-6 UEFC-1 UEFC-3 UEMP-5 UEFG-1 UEMG-3 UEFP-2 SEMP-5 SEMP-2	UEFP-4 UEMP-5 DHFP-8 DHFP-9 DHFP-10 DHMP-12 SEMP-5 SEMP-2	DHFC-2 DHMC-5 DHMC-6 DHFC-7 DHFP-8 DHFP-10 DHMP-12 DHFG-2 UEFC-1 UEFC-3 UEMC-6 UEFP-4 UEFP-2 SEMP-5 UEMP-5 UEFG-1 UEMG-3 SEFC-1 SEFC-3 SEFC-4 SEMP-2 SEFP-6
Food consumption	DHMC-1 DHFC-2 DHMC-3 DHMC-4 DHMC-5 DHMC-6 DHFC-7 DHMC-11 DHFP-9 DHFP-10 DHMP-12 DHMG-4 UEFC-1 UEFP-2 UEFG-1 SEFC-1 SEMP-2	DHFP-8 DHMG-4	DHMC-1 DHMC-5 DHMC-6 DHFC-7 DHFG-2 DHMG-4 UEFC-1 UEMC-6 UEFC-3 UEFP-2 UEFP-4 UEMG-3 SEFC-1 SEFC-3 SEMP-2

Where blue is used for middle income households, orange for poorer households and green for focus groups.

So far we have focused on coding of attribution tags. Having digested this headline information (how many respondents cited the project positively and negatively, as well as explicitly, implicitly or not at all, and in relation to which outcome domains?) readers will quickly want more specific information about what these **drivers** are. This presents a third challenge for visualisation and synthesis. The QuIP tackles it through thematic and mostly inductive (*ex post*) coding of drivers of change into clusters, some explicitly linked to the project and others not. Our visualisations here show the number of citations of each driver across all domains (citation count) as well as disaggregation by domain which shows how many respondents cited the driver in each domain (respondent count).

Figure 5: Example presentation of negative drivers of change



Metric: Each colour uses respondent counts – adding up to total citation counts

More disaggregated data is presented in a table in an appendix, showing unique household IDs by domain (as columns) and driver (as rows).

The same approach of considering both sets of numbers is used for presenting outcome data. In some cases outcome domains are too broad to be very insightful, so the QuIP coding system includes an option to tag each causal claim with an additional and more detailed thematic **outcome** code. Additional outcome codes are generated inductively, derived from the project's theory of change, or a combination of both.

Figure 6: Example presentation of negative outcomes

Metric: Respondent counts (citation counts) [citation intensity counts]

In this example counts of different coded responses are presented by individual respondent and focus group separately (respondent counts), supported separately by the total number of times it was repeated by them in brackets (citation counts), followed by the mean number of citations per respondent (citation intensity count) in square brackets. 18+4 (29) [1.3]' is used as shorthand for 18 individuals and 4 focus groups making a specific causal claim, repeated a total of 29 times across all outcome domains, resulting in a mean of 1.3 citations per respondent.

To elaborate further on this point, consider the difference between the number of cases (out of 28) that made a specific causal claim with respect to any outcome domain, 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 in any domain. 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 domains), or from all 24 respondents and with reference to just one domain. The quality of this indicator will also depend upon whether the 24 included any focus groups or just individual respondents.

It is consistent with the exploratory and blindfolded nature of the QuIP to start by documenting outcomes and then drill back to presentation of drivers of those outcomes. But in some cases, users of the evidence may also be interested in the data being presented the other way around: i.e. starting with specified project actions (as drivers of change) and summarising the evidence (in the form of frequency counts) on how many and which outcomes respondents connected them to.

Figure 7: Extract from example of positive outcomes and associated drivers of change

	Outcomes				
Driver of Change	Increased income	Clean drinking water	Improved food production/ consumption	Healthier livestock	Expenditure decreased
Cash for work programme	49		6		
Rehabilitation of tater sources		45	8		16
Livestock feed project	2		9	29	5
Rain/ recovering from drought	6		12	14	1
Alternative income (shop)	5		2		
Increased price of livestock	11		2		

Metric: Citation counts

It is possible to filter the evidence by drivers or by outcomes, however it is important to highlight the risk that readers will pay too much attention to evidence of causation linked to programme-related drivers (which is what they most want to see) relative to other (incidental) drivers of change, including the influence of other development interventions affecting the same intended beneficiaries. A useful term for bias towards drivers of change emanating from our own actions compared to other drivers is *solipsism*, and an important part of the role of the QuIP as a reality check is to help protect development agencies from this bias. To this end all driver/outcome relationships are presented in tabular form, albeit sometimes clustered by attribution type.

Having worked through the main course of data presentation, we now arrive at the potential 'show stopper'. The analyst has the option to build up primary, secondary and tertiary outcome codes, thereby building up a causal chain. This comprises flow diagrams that indicate multiple causal links from higher-level outcomes back to intermediate incomes and underlying drivers of change. The importance of these causal pathways across populations can be illustrated with respondent and citation counts. For a three-step pathway these are as follows: the causal driver (X), the causal claim linking the driver to a specified primary outcome ($X->Y_1$), the primary outcome (Y_1), the causal claim linking the primary outcome to a secondary outcome ($Y_1->Y_2$), and the secondary outcome (Y_2). Diagrams can reveal this data visually, for example by altering the thickness of arrows.

Figure 8: Example presentation of drivers and outcomes

Metric: Citation counts

Here, as with all other numerical representation of findings, it is important to complement the use of frequency counts as proxy indicators of the importance of different causal processes with discussion and quotation of the underlying text. 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 indicators of statistical significance. There are two reasons for this: first, they are not usually based on representative samples; and second, the 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.

In this sense, and returning to the different ways of interpreting the *qual/quant* divide; whilst the QuIP certainly entails processes of codification, and may thus appear to retain features of a quantitative methodological approach, it remains strictly a qualitative and interpretive tool.

A last thought concerns the potential of the QuIP to furnish evidence on complex and heterogeneous causal processes. Where an outcome arises only through the conjunction of two or more causal drivers then coding may reflect this (i.e. ascribing separate codes to different packages of causes cited by respondents). However, more generally it is part of the art of the analyst 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 also made possible by timely and sufficient rainfall. Likewise some causal pathways may be restricted to sub-groups of the full population of intended beneficiaries (e.g. women, those living near a road, or in just one village), and it is again part of the skill of the analyst to identify this, or raise it as a possibility.

Conclusion

While the QuIP is a qualitative approach to impact evaluation it can not only be used in conjunction with quantitative methods, but involves processes of codification and counting that give it at least a partially quantitative flavour. 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. 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. 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 pickings on offer. Nor does it undermine the qualitative and interpretive philosophy underpinning the QuIP as an impact evaluation method.

Appendix: QuIP glossary

Attribution. Evidence that an action (X) of a named organisation or project is contributing to change in an *outcome* (Y) in the presence of other *drivers of change* (Z).

Attribution code. A code that indicates whether a *causal claim* (a) is having either a positive, negative or neutral effect on a specified *outcome*, and (b) explicitly identifies a selected organisation as the *driver of change*, is implicitly consistent with its *theory of change*, or is unrelated/incidental to it actions.

Causal claim. A proposition that a specified *outcome* (Y) was a direct consequence of a specified action (X) or (Z).

Causal driver. See *driver of change*.

Citation count. The total number of times a particular driver of change, causal claim or outcome is coded, including multiple coding for the same respondent across more than one domain (unlike the *respondent count*).

Citation intensity. The mean number of citations of a coded driver, causal claim or outcome per respondent. Hence, if C is the citation count, R is the respondent count and I is the Citation intensity then I=C/R.

Commissioner. The organisation contracting a QuIP study, and the primary user of the evidence to be collected. Responsibility rests with them to decide what sort of evidence they want, as well as when, where, how and why to collect it.

Credible cause. A *driver of change (*X), credibly causes *outcome* (Y) in a particular context if (i) is strong evidence that X and Y happened, (ii) several stakeholders independently assert that X was a cause of Y, with minimal prompting, (iii) there is no more credible counter-explanation for why they might have said this, (iv) their account of how X caused Y is consistent with a plausible *theory of change*.

Domain. A field or category of outcomes, agreed in advance with the commissioner and used to structure interviews and focus group discussions. Most studies address a set or group of domains

that are consistent with a *theory of change*. For example, they may refer to different aspects of the well-being of individual intended beneficiaries.

Driver of change. An action or state (X or Z) behind outcomes (Y). These are generally self- reported by respondents, in answer to questions like *'why did that happen?' or 'what was the reason for that?'* This term is synonymous with *causal driver*. Thematic coding is used to group similar drivers together into groups or clusters.

Intended beneficiary. Those people that a specified organisation is aiming to benefit, by achieving *outcomes* specified in its *theory of change*. In the case of capacity building projects the intended beneficiaries may be organisations or associations of people.

Impact. Evidence that a specified project credibly caused a specified set of outcomes. In some cases the term impact may refer specifically to final or *tertiary outcomes*.

Outcomes. Changes (positive or negative) reported by respondents, often in the answer to the question *'during the last [specified time period] has anything changed in relation to [domain of wellbeing]?'* Since outcomes can also become drivers of change, we code primary, secondary and tertiary outcomes if required. For example, X may lead to Y_1 leading to Y_2 leading to Y_3. In this case Y_1 and Y_2 are both drivers of change and outcomes (primary and secondary). These intermediate outcomes may also be referred to by others as outputs or results, but in QuIP studies these terms are generally avoided.

Project. A specified set of activities, intervention, investments over a given period of time aimed at achieving a specified set of intended outcomes for a specified group of intended beneficiaries. This is the object of a specified QuIP study, and it is the responsibility of the commissioner to define it, as well as the theory of change behind it, as precisely as possible. Others may refer to the project as a 'treatment' but in QuIP studies this term is generally avoided.

Respondents. These are the main source of causal claims, linking drivers of change (including but not limited to project activities) to outcomes, both intended and unintended. Respondents are usually a sample of intended beneficiaries, and data is collected form them through a mix of semi-structured interviews and focus group discussions.

Respondent count. The number of respondents (usually counting a focus group as a single respondent) for which a particular *driver of change*, *causal claim* or *outcome* is coded at least once. If the same driver, claim or outcome is coded across more than one domain for the same respondent then the count remains one (in contrast to the *citation count*). Respondent counts are proxy indicators of the importance of a coded finding because they indicate how widely it was independently reported. However, as QuIP studies are usually not based on a representative sample of intended beneficiaries these counts are a weak indicator of overall significance. They are better thoughts of in Bayesian terms. For example, given prior expectations that the project is a driver of a specified positive outcome, how frequently would you expect it to be mentioned explicitly? Or how much would you modify your prior view if it was not mentioned at all?

Theory of change. The causal processes by which the commissioner of QuIP study expects a specified project to achieve intended outcomes and impact. Not all causal drivers originate with the project. Theories of change also identify incidental drivers of change and may also assess the risks associated with their occurrence or non-occurrence.