QuIP Data Analysis: Note on amendment to coding of QuIP data

This brief explains the revised approach to coding QuIP data; why the change was made, and what difference this will make to coding and analysis. This note is written for experienced analysts who were trained in QuIP coding prior to <u>May 2019</u> - it is not a briefing note for new analysts.

Previous approach

Stories of change collected in QuIP studies have hitherto been coded using the QuIP 'triple' coding approach; coding segments of the data that make <u>causal claims</u> as either:

- <u>Drivers of change</u> - based on inductive classification of the reasons behind any change or outcome; or,

- <u>Outcomes</u> - based on inductive classification, and allowing for up to three linked outcomes to be classified - one outcome leading to another, and then to another outcome;

This is in addition to an <u>attribution</u> code - deductive coding based on predetermined codes that provide an initial indication of the strength of the attribution claim.

Coding is done in an Excel sheet, using columns to denote the position of a code in a causal claim. Hitherto, any tag used in the first column was considered a driver (using the pre-fix D), and any tags following were denoted outcomes (using the pre-fixes OP or ON, depending on whether they were positive or negative changes). An example is given below:

Column 1 (Driver)	Column 2 (Outcome 1)	Column 3 (Outcome 2)	Column 4 (Outcome 3)
D1: Access to hybrid seeds from X	OP1: Used hybrid seeds	OP2: Increased production	OP3: Increased income

Why change?

The change has been introduced to give data analysts more flexibility in the way that they code, and to avoid some potential pitfalls with fixing Driver/ Outcome categories. By fixing a description (tag) as either a driver or an outcome in an analyst's tagging key, this sometimes meant duplicating a tag in both the driver and outcome lists. The example below demonstrates how a tag could be either a driver or outcome, depending on the type of response being coded. Some narrative extracts provide a full story of change, with root cause. Others provide only part of this story, not going back as far, e.g.

Respondent 1 says:

Column 1 (Driver)	Column 2 (Outcome 1)	Column 3 (Outcome 2)	Column 4 (Outcome 3)
D1: Access to hybrid seeds from X	OP1: Used hybrid seeds	OP2: Increased production	OP3: Increased income

Respondent 2 says:

Column 1 (Driver)	Column 2 (Outcome 1)	Column 3 (Outcome 2)	Column 4 (Outcome 3)
D2: Use hybrid Seeds	OP2: Increased production	OP3: Increased income	

In this example, the 'Used hybrid seeds' tag is the first point of change for Respondent 2, and therefore a Driver; they haven't mentioned the source of the hybrid seeds. For Respondent 1, this same tag is the second point of change, resulting from 'Access to hybrid seeds'. This means a different tag is used to describe the same story across two different respondents; D2 and OP1.

This does not cause problems in relation to the causal claim visualisations produced in the dashboard since these only count connections between columns, or nodes, rather than counting the variables within the columns. However, when looking at definitions of and counts of drivers and outcomes, it can cause confusion and duplication.

New approach

a) No distinction between drivers and outcomes

The new system of coding removes the need to define any tag as a driver or an outcome - this will now be calculated automatically depending on its position in the columns.

- A tag <u>followed by any other tags</u> in the same row of text will automatically be considered a source, or Driver of change. It is considered to have a causal relationship with the tags following it.
- Any tag at the <u>end</u> of a row of tags, with nothing following it, will automatically be considered a target, or Outcome.

Of course, the distinction between the labels driver and outcome remains blurred; a tag used in the middle of a causal chain is both a driver and an outcome, but for the purposes of analysis we are now linking these definitions to the idea of 'source' and 'target'. This is the logic used in the software for producing causal chains, based on Sankey diagrams which distinguish between source nodes and target nodes at the end of the connections. This means that most tags will now fall into the 'driver' typology, with only final positions defined as 'outcomes'. This position will be whichever was the last to be completed, which could be column 2, 3 or 4.

The main benefit of this change is that the analyst does not have to decide whether any tag should be consistently used as a driver or an outcome, it can be either or both, reflecting the stories as they are told by each respondent. The resulting counts will be a more accurate reflection of the column position that the main stories fall into.

b) No distinction between positive and negative tags

We have also removed the need to code tags as either positive or negative. This will be automatically inferred from the attribution code used in the row, which already refers to the causal statement as either positive or negative. The main benefit of this change is that the analyst can use the same tag even if one respondent considers it to be positive, whilst another considers it to be negative. This avoids the need to create different tags for each scenario.

New tagging system:

This update means changes to the tagging key used across all QuIP studies. The tags used as shorthand for the descriptions of drivers or outcomes will now be purely numerical. There will be one list of tags, compared to the three lists which existed before, for drivers, positive outcomes and negative outcomes. An example is below, where you will note that there is no D or O pre-fix, and no positive or negative code used.

Tag	Description
1	Hybrid seeds given by Org X
2	Used hybrid seeds
3	Production increased
4	Income increased
5	Family nutrition improved

Some analysts may prefer to use three-digit numbers to help with thematic clustering of tags, e.g. 101, 102, 201, 202. Analysts are free to use whatever numerical codes they wish, there is no restriction or guidance on this.

This also necessitates a change to the attribution codes, moving away from numerical codes to alphabetical codes. The new attribution table is below, with the previous codes in brackets.

Attribution codes for confirmatory analysis of project impact

Positive		**Negative**
PE (1)	Change explicitly attributed to project and project-linked activities	**NE** (2)
PI (3)	Stories confirming/questioning a mechanism by which the project aims to be achieving impact, but with no explicit reference to the project	**NI** (4)
PO (5)	Change attributed to any other forces that are not related to activities included in the commissioning agent's theory of change	**NO** (6)
PN (7)	Change not attributed to any specific cause	**NN** (8)
O (9)	Other	

An example of how text would be coded in the spreadsheet is below:

	Position 1	Position 2	Position 3	Position 4	Attribution
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.	1	5	3	10	PE
Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.	1	5	3		PE
Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.	6	7			NO

The change to the outputs in the analysis dashboard are minimal and relate mainly to the pages which count Drivers and Outcomes. These will still serve the same function and show data in the same way, but you should be aware of how the distinction has been made between them since this is now done automatically - and will lead to a higher count of drivers and a lower count of outcomes compared to the previous approach. There is no change to counts used in the visualisation of causal chains.

Importantly, we can now count tags depending on their position, allowing the analyst to see whether a tag has been used predominantly in an early, mid or end position across the dataset. This is found in chapter 9 of the dashboard, under Overall Counts By Position, both as a table and a bar chart.

	Position 1	Position 2	Position 3	Position 4
Description	**Citation Count**			
Able to hire/pay workers for farm		1	7	
Able to pay school fees/uniform		2	19	
Access to seeds	3			
Access to cuttings	6			
Access to treated water	10	4		
Agricultural training and advice	52			
Able to hire/pay workers for farm		1	7	
Able to pay school fees/uniform/		2	19	
Improved health			8	21

The Causal Chain Calculator table under chapter 6 of the dashboard now also shows tags separately if they have appeared in different positions.

This change to coding and analysis is true to the QuIP spirit of continuous improvement, and we hope that it will help to improve the flexibility and accuracy of the approach.

For more information:

info@bathsdr.org

www.bathsdr.org

BSDR · Bath Social & Development Research Ltd