Headlight
CONSULTING SERVICES

Qualitative Comparative Analysis Methods Memo

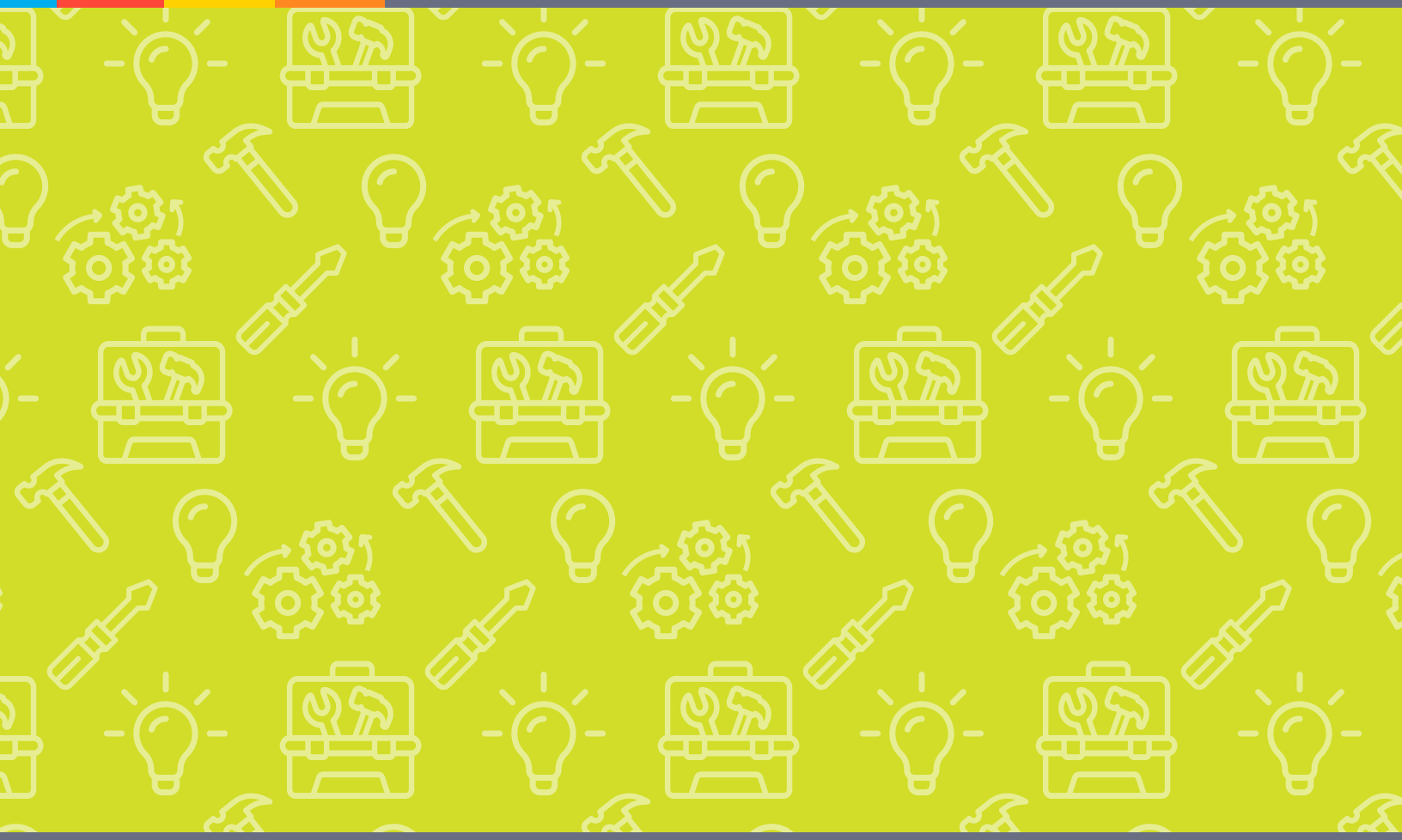


Table of Contents

Acknowledgments

During the process of writing this Methods Memo, we have been struck by how inquisitive and interested our colleagues have been about Qualitative Comparative Analysis (QCA) as a method and the increased interest in the field in rigorous qualitative approaches. We have also had the extraordinary privilege to build upon the existing resources and conceptualization of the method established by Charles Ragin and later iterated on by many other social scientists. We are deeply appreciative of our colleagues at the USAID/Ethiopia Mission who delved into exploring how QCA could help them better understand Crisis Modifiers at the nexus of development and humanitarian programming with us in 2022. And to other colleagues and evaluators who offered their perspectives, experiences, and learnings during discussions about this method over the past few years.

While this Methods Memo can never be inclusive of all information on the method, we hope that this is a good place to start for all those looking for practical implementation guidance.

We would like to thank the team responsible for developing the content within this Methods Memo, including Rebecca Herrington, Chelsie Kuhn, and Alison Harrell. If there are further questions, please feel free to reach out to us at info@headlightconsultingservices.com.

Authors

Rebecca Herrington

CEO, Headlight Consulting Services

Rebecca Herrington is an evaluation, learning, and adaptation technical specialist with 16 years' experience focused on evidence-based strategy design, systems thinking, and collective action. She is an expert and thought leader in complexity-aware collaborating, learning, adapting, monitoring, and evaluation (CLAME) and in qualitative methods and rigor, including but not limited to Developmental Evaluation, Process Tracing, Outcome Harvesting, Qualitative Comparative Analysis, and Positive Deviance. Ms. Herrington's work focuses on international development and conflict, specifically complexity programming, with a keen ability to facilitate evidence to action and provide strategic communications. Ms. Herrington has field and academic experience throughout Central America, MENA, and East Africa, with over 5 years living and working abroad in rural, conflict-affected, and transitional communities. Her experience is further grounded in a Masters of Sustainable International Development and a Masters in Coexistence and Conflict Management, both from Brandeis University. Some highlights from her accomplishments include, but are not limited to, developing the theory of change and MEL Plan with continued learning advisory support to the USAID Digital Strategy, building an evidence base CLAME system for GSMA's Mobile for Development programs, supporting integrated Systems and Evidence Based Theory of Change and Learning Agendas for USAID/Ethiopia across their emergency response, disaster risk management, and resilience portfolios, conducting Learning Reviews for GSMA, USAID/Bosnia and Herzegovina, USAID/Ethiopia, and designing a meta-analysis to support evidentiary revisions of the USAID/Resilience and Food Security Institutional Architecture Framework, and managing the successful growth of Headlight's first five years.

Alison Harrell

CLAME Specialist, Headlight Consulting Services

Alison is an experienced professional with 9 years of experience in international development and public service, and 7 of those years were spent honing and applying skills in monitoring, evaluation, research, and learning. Most recently, Alison worked for 3.5 years as a contractor supporting the USAID Digital Finance team by leading the team's monitoring, evaluation, and learning portfolio, managing 4 research projects on digital financial services, and enhancing the team's knowledge management and decision-making processes. She has field and academic experiences in Eastern and Southern Africa and Latin America regions. Alison has additional work experience at the U.S. Department of State, AmeriCorps, America's Unofficial Ambassadors, and the Ethiopian Community Development Council. Alison received a Master of International Public Affairs at the University of Wisconsin - Madison and a Bachelor of Science in Community, Environment, and Development with a focus in International Development at the Pennsylvania State University. When first joining Headlight, Alison served as the DE Deputy Administrator implementing a Developmental Evaluation (DE) with the USAID/Ethiopia Disaster Risk Management and Effective Emergency Response teams. Currently, Alison is serving as the DE Deputy Administrator for a 5-year Mission-wide DE with USAID/Ethiopia, focused on improving learning, adaptation, and evidence-based decision-making across the Mission.

Chelsie Kuhn

CLAME Specialist, Headlight Consulting Services

Chelsie Kuhn is a dedicated professional with 8 years of experience in supporting efficient, effective, and evidence-based operations both with NGOs and in academia. She brings a strong background in managing large groups of diverse stakeholders and a data-driven mindset to project management. Prior to joining Headlight, Ms. Kuhn was responsible for assisting more than 20 offices at a large 4-year public university with assessing program outcomes, developing and coordinating strategic plans, meeting annual reporting deadlines, and working toward continuous improvement through collaborative learning efforts. Since joining Headlight, she has contributed to qualitative data analysis and report writing to develop a Learning Review for USAID/Bosnia and Herzegovina and for USAID/Ethiopia's Disaster Risk Management (DRM) portfolio, drafted MEL Frameworks and designed MEL tools for GSMA's Mobile for Development (M4D) programs, and distilled how-to oriented learnings for MEL practitioners of all levels. As a result of her work co-leading a multi-faceted qualitative coding and data analysis team effort to produce the first Learning Review of disaster risk management and resilience implementation in Ethiopia, various stakeholders have been able to identify and plan for adaptations to improve their work and best serve the most vulnerable Ethiopians. Following this Learning Review, she worked to co-implement a QCA evaluation with Outcome Harvesting embedded into it to explore USAID/Ethiopia's use of Crisis Modifiers (flexible funding awards) in development programs to address emergency humanitarian needs. As a result of that effort, implementers have been able to use the findings to make improvements in their work. She is currently working on providing technical direction and additional support to a 5-year Mission-wide Developmental Evaluation for USAID/Ethiopia. Her experience is grounded with both a Master's and Bachelor's degree in Conflict Analysis and Resolution from George Mason University.

Acronyms

CLAME	Collaborating, Learning, Adapting, Monitoring, and Evaluation
CM	Crisis Modifiers
DE	Developmental Evaluation
DRM	Disaster Risk Management
EER	Effective Emergency Response
FCR	Findings, Conclusions, and Recommendations
fsQCA	Fuzzy Set Qualitative Comparative Analysis
LOE	Level of Effort
MEL	Monitoring, Evaluation, and Learning
OH	Outcome Harvesting
PD	Positive Deviance
QCA	Qualitative Comparative Analysis
RCT	Randomized Control Trials
RF	Rapid Feedback Experiments
SAGE	Strategic Advisory Group for Emergencies
USAID	United States Agency for International Development

Glossary

Term	Definition
Case	a singular activity or intervention that the evaluator wants to explore further in order to be able to make comparisons and inferences about why change happens in some situations but not others. A case should be definitive with a clear who, what, where, and when.
Causal Mechanism	the step-by-step process for how a change occurs. A causal mechanism is the series of actions by different actors that causes X intervention to lead to Y outcome.
Complexity	an indicator of the inter-relationships and connections within a system that affect the way people and programs relate to each other and influence their ability to achieve desired objectives. An increase in inter-relationships and/or connections increased complexity.
Factor	conditions and details, internal or external, that describe and enable comparison of cases. The factors' presence or absence should contribute to understanding the results of the selected cases.
Method	structured processes and/or frameworks that help to best answer evaluative questions or capture the information needed.
Outcome	a change in the behavior, relationships, activities, policies, or practices of an individual, group, community, organization, or institution.[1]
Sampling	the process of identifying from whom the evaluators will collect data and how; it is important that the chosen sampling method is appropriate for the evaluation questions that have been identified, the type(s) of information the evaluators need, and any limitations they might face in data collection.
Sampling Saturation	the point when incoming data produces little to no new information (Guest et al., 2006; Guest and MacQueen, 2008); There is a wide range of existing research on this topic, and most sources agree that at least six interviews of a homogeneous group (as defined by the evaluative effort sampling structure) will cover 70% or more of the findings that will emerge from further data collection (Guest et al., 2006). According to Guest et al., 12 interviews will increase that coverage to 92%.
Systems	a group of interrelated parts that come together to form a more complex, functioning whole that serves a specific purpose.
Triangulation	when three or more sources confirm that something has occurred; triangulation allows for quality assurance around the rigor of data and findings.
Qualitative Data	data collected using data collection methods, such as interviews, focus groups, observation, and key informant interviews. Qualitative data can provide an understanding of social situations and interactions, as well as people's values, perceptions, motivations, and reactions. Qualitative data are generally expressed in narrative form, pictures, or objects (i.e., not numerically).
Quantitative Data	data collected using quantitative methods, such as surveys. Quantitative data are measured on a numerical scale and can be analyzed using statistical methods.

1 Wilson-Grau, R. and Britt, H., 2012. Outcome Harvesting. Ford Foundation. Available at <https://outcomeharvesting.net/outcome-harvesting-brief/>

Introduction

About us

Headlight Consulting Services (Headlight) is a women-owned, small business established in 2018 to facilitate data-driven decision-making through systematic design support; in-depth monitoring, evaluation, and learning technical services; and facilitated organizational change processes. Headlight helps clients design and integrate evidence-based collaboration and learning efforts by focusing on the structures and systems that will enable sustainable solutions. Headlight's clients include the United States Agency for International Development (USAID) Missions, USAID Washington, private sector actors, and more. In all of its efforts, Headlight takes a local-first approach, striving to sustainably invest in local talent while providing exemplary quality assurance from Headquarters to ensure cost-effective and high-quality services.

Headlight Consulting Services seeks to strengthen international development outcomes in sustainable and locally-owned ways. One of our contributions towards this systems-level objective is the provision of use-focused tools, facilitated on-the-job training, and complementary capacity-strengthening supports in Collaborating, Learning, Adapting, Monitoring, and Evaluation (CLAME). Our hope is that with more access to easy-to-use tools, training, and mentorship in CLAME, practitioners can improve their data-driven decision-making practices, thereby improving development interventions and their outcomes.

This Methods Memo is the second in a series of more robust products intended to provide how-to guidance for professionals of all levels to implement stronger CLAME practices. This particular memo's goal is to provide evaluators with practical guidance for implementing the Qualitative Comparative Analysis evaluation method. Its contents may also be relevant to organizations that are looking to establish and improve their qualitative methods considerations and practices. The sub-sections below are designed as use-focused modules intended to enable practitioners to walk step-by-step through the whole process or dive directly into a particular component of the method they want to strengthen. We hope that this note will help inspire evaluators, project leaders, and donors alike to start or continue applying the Qualitative Comparative Analysis method where appropriate and to fuel further innovation, rigor, and adaptation in their work.

MODULE 1:

What is Qualitative Comparative Analysis?

DEFINITION

Qualitative Comparative Analysis (QCA) is an evaluation and research method that, "enables the analysis of multiple cases in complex situations, and can help explain why change happens in some cases but not others," (Simister and Scholz, 2017). The analysis is based on identifying a set of cases (individual activities or interventions) that have some common alignment and analyzing each case according to a set of factors to establish a cross-case comparison.

The most unique aspect of QCA is the comparison it allows among (and within) cases. Rather than compiling a set of separate case studies and observationally noting the similarities and differences, QCA offers a more rigorous approach to analyzing similarities, differences, influencing factors, etc. across cases and identifying the factors critical for success or achieving desired outcomes. A few other key aspects of this method include:

- The approach is based on assumptions that a) change often results from different combinations of factors and b) different combinations of factors can produce similar changes (Ragin, 1984).
- QCA is best suited for an intermediate number of cases (typically between 10-50), demonstrating its utility when the cases of interest are too few for conventional statistical analysis and too large for another type of qualitative deep-dive case study approach like Positive Deviance (Simister and Scholz, 2017).[2]
- The original application of QCA analyzed factors against Likert scales, enabling a level of quantitative analysis; however, its evolved use in evaluation often leverages a range of analysis methods dependent on the factors chosen to align with comparative information needs, such as Thematic, Process, Time-based/Sequential, and Narrative Analysis.

ORIGINATION OF METHOD AND USES TO DATE

Qualitative Comparative Analysis originated as a new method in the social sciences in the late 1980s to help researchers compare cases and parse through multiple variables to explain how cases were similar to and different from one another (Marx, Rihoux, and Ragin, 2013). Since Charles Ragin published the source material in *The Comparative Method* in 1987, social scientists, including sociologists, anthropologists, and political scientists, have leveraged this approach to compare cases of revolutions, social movements, unionization, and more to improve the world's understanding of what change(s) occur depending on circumstances, context, and other influencing factors (Marx, Rihoux, and Ragin, 2013). To build upon previous areas of controversy, Ragin, Dirk Berg-Scholosser, and others worked to clarify how to include cases where variables are not easily sorted into binary categories; establish new strategies for selecting conditions and variables; develop ways to examine

2 Please see the Designing the Evaluation section in Module 3 for more in-depth guidance on determining the appropriate number of cases and case selection.

elements of Process, Sequential, and Temporal Analysis; address consistency and coverage to explore causal mechanisms and combinations; establish other goodness-of-fit tests to assess how well empirical models fit and account for measurement errors; and, explore how to layer QCA with other methods and techniques (Marx, Rihoux, and Ragin, 2013). For information on each of these evolutions, please reference the cited article *"The origins, development, and application of Qualitative Comparative Analysis: the first 25 years."*

HEADLIGHT'S VARIATION ON THE METHOD

Acknowledging that QCA is not a new method in the social science field, Headlight staff have learned a lot about this method's application in international development and associated fields. Headlight's particular focus has been on practical right-sizing of the method to strike the balance between structure and rigor[3], contextualizing application based on clients' needs, and promoting evaluation use[4] for improved evidence-driven decision-making. Applying this approach to QCA, Headlight recommends a couple of variations on the method, noting that design should stem from the client's overarching evaluation question(s) and purpose.

First, Headlight focuses on case selection to be a smaller set of 6-12 cases instead of 10-50 cases. This is a tradeoff considering the more qualitative approach where sampling strategies and sampling saturation standards ensure rigor and saturation in regards to factors seeking to be compared while also appropriately managing timelines and level of effort (LOE) required for implementation. For example, if 12 interviews are going to be conducted for maximum saturation coverage per case, then 12 cases would yield 144 interviews that then need to be processed, coded, and analyzed. While researchers may be able to manage more, we would advise evaluation designers to take the sampling needs into consideration alongside the goals of their effort.

Second, most guidance on implementing QCA relies on using computer software to assign numerical values to each factor, conduct quantitative analysis, and present the set of factors that achieved the desired result. Instead of relying on quantitative assignments to provide rigor, at Headlight, the evaluation team combines fit-for-purpose sampling strategies, sampling saturation quotas, and a variety of analytical methods to compare the factors. This approach best enables the comparison of **more nuanced and complex qualitative aspects of cases**, prevents prescriptive treatment of factors before data collection, and tailors analysis of individual factors to answer specific evaluation questions.

3 Rigorous evidence is conducted according to the highest standards of the methodology that is best suited to the specific nature of the study– all methods used in the study (e.g., evaluation, sampling, data collection, and analysis). Evidence gathering must adequately address issues of both internal and external validity and ensure accurate reporting of results while protecting sensitive subject data. High standards of rigor ensure the integrity of the evidence generation process and results, ensuring the evidence is trustworthy. Highly rigorous evidence often covers multiple contexts and meets sufficient saturation and/or statistical significance to be generalizable.

4 Use-focused evidence is defined by its intentional design to meet the audience's information gap or needs and facilitate subsequent uptake and application of both findings and recommendations through updated understanding of topics covered by the evidence and implementation of adaptations to current and future actions by the intended users. As such, use-focused evidence is best measured post-dissemination by assessing actual use cases. However, several predictive metrics can be applied to assess prepared evidence for the likelihood of its use, including but not limited to the timeliness of delivery (i.e., was the evidence delivered before a key decision point), the appropriate use of data visualization that clearly presents key takeaways, a report structure that purposefully guides the intended user(s) from their current knowledge and understanding through the new learnings, recommendations that are contextualized, feasible, and actionable with specifics of who should do what by when and how, and an evidence dissemination plan that takes into account intended user decision points and processing perspectives.

To take this altered approach to QCA, an evaluator should ask questions like –

- What cases and factors are most essential to answer the evaluation questions?
- What type of information is needed to assess that factor and what combination of interviewees is needed to understand the factor in full?
- What type(s) of analysis methods will provide the most clear and relevant findings for each factor?

This approach allows for analysis of a greater variation of information, not only factors that fit well within a numerical assignment (for specific examples, please reference Table 2 in Module 3). Especially in complex operating environments, applying additional qualitative analysis methods allows an evaluator to more fully assess and understand the "why", "how", and enabling and inhibiting environments for the cases, as well as gather the information that can be used by stakeholders to implement adaptations based on the evidence.

What is QCA?

- QCA compares similar cases across a set of factors and analyzes each using best-fit methods per factor.
- The result of applying QCA is a detailed analysis of the influence of the different factors on each case and broader takeaways from cross-case comparison– not only what factors most often lead to success, but findings on how certain factors interact with each other, environmental or relational dependencies for change, what inhibiting environment conditions should be mitigated against, etc.
- The resulting information can be used to adapt implementation and propel learning for continuous improvement.

What is QCA not?

- **QCA is not a collection of case studies.** While the method does use a case-based approach, gathering and summarizing seemingly similar case studies does not qualify as QCA. The specific focus of QCA on factors and rigorous Comparative Analysis goes beyond a simple summary of cases or observation of similarities and differences across intervention anecdotes.
- **QCA is not a quantitative approach masquerading as a qualitative method.** This is often a misconception of the method and can lead to misapplication. While there are aspects of the method that may rely on quantitative analysis, depending on how the evaluator structures chosen factors, QCA is first and foremost a qualitative method.
- **QCA is not suited to be a secondary output of another primary method.** QCA does not naturally embed within other evaluation methods because of the level of complexity the analysis requires. QCA should be relied upon as a primary evaluation method and secondary methods can be embedded as factors as needed (see this example of embedding Outcome Harvesting in Module 6).

The next Module, "How Does an Evaluator Determine When to Use QCA?", will compare and contrast QCA with other similar methods to enable use-focused evaluation design, provide guidance on what types of evaluation questions may be a good fit, and explain what to expect when implementing QCA.

MODULE 2:

How Does an Evaluator Determine When to Use QCA?

This module will explain how to assess if QCA is the best-fit method to answer identified evaluation questions and meet the overarching evaluation objective. Evaluators will learn how to identify when evaluation questions are well-suited for QCA, when to use QCA versus other methods, what the limitations of the method are, and what to expect before implementing the method.

To begin, an evaluator should assess what the information needs are for the particular evaluation, typically specified through evaluation questions. The evaluation questions should drive the selection of the evaluation method, based on what type of method will provide the best structure to obtain the information needed to answer the evaluation questions. Table 1 presents several illustrative examples of the types of information needs that are best suited to QCA, associated evaluation question examples, and an explanation of the relevance and rationale of why QCA would be an appropriate method to support the collection and analysis of the desired information.

Types of Relevant Evaluation Questions

Category of Information Need	Types of Evaluation Questions	Relevance/Rationale for QCA
Comparison of Intervention Sites	At which intervention site(s) was XYZ factor particularly effective or not effective? What were the greatest differences in factors that influenced X intervention in XYZ intervention sites? How did those differences affect the programs' results?	The set of cases includes various intervention sites, some of which had XYZ factor present and some of which did not. QCA will enable comparison across the cases to understand what role XYZ factor played in each case and how this factor interacted with other factors of interest.
Understanding of Different Contexts	How do the contextual differences in XYZ intervention sites affect the communities' level of engagement? How does the population size of the intervention area affect implementation?	QCA is a great fit when looking to determine how different contexts (culture, societal norms, environment, weather patterns, history, etc.) affect implementation. All interventions are influenced by their context, which will differ even from village to village. QCA provides a way to break down key contextual factors and compare and contrast **them to** determine their effect on intervention(s) efficacy and implementation strategies.
Identification of Influential Inhibitors and Enablers	What enabling and inhibiting factors most influenced the results of XYZ intervention across the identified cases?	The presence of specific enabling and inhibiting factors can be compared across QCA cases to assess how they did or did not influence each case, which influencing factors are most common, and potential mitigating strategies implemented in successful cases.

EXPLORING THE RIGHT METHOD FIT

When designing an evaluation, it is important to not just think immediately of the data collection approaches the evaluator might take (e.g., mixed methods including both quantitative and qualitative data) and sampling strategies (e.g., purposive, snowball, stratified random, cluster, etc.), but to first find the evaluation method that will best help structure the overall evaluative effort to appropriately answer the identified evaluation questions. QCA is not a one-size-fits-all evaluation method and should only be chosen if it will meet the desired aims of the evaluation. The section below will help explain when QCA is most useful, identify questions to help evaluators confirm that QCA is the right approach, identify limitations to QCA, and recommend other methods and approaches for when QCA is either not enough on its own or does not fit the evaluator's needs.

QCA is best used as a **summative method** after the implementation of an activity or series of related interventions has been completed. There is often not enough information to compare while interventions are still happening (e.g., as a formative evaluation), it may not be clear what the most advantageous factors to look at are, and there may be too many gaps in the data availability. QCA is also generally **goal-oriented** (as opposed to goal-free), since the method is designed to compare and contrast a number of factors and the influence of the context on individual activities/interventions that are believed to have some common alignment. QCA also **strikes the balance between evaluation for accountability and evaluation for learning** in that seeking to understand change, or lack thereof, can be used to propel future efforts with lessons learned while also serving as an assessment of the results from what has been implemented to date.

When thinking about QCA as compared to other methods, there are a few orienting questions to help evaluators choose the right approach:

?
- Are you looking to analyze similar cases to understand how a variable influences change?
- Do you want to understand more about how similar factors/variables can lead to different outcomes?
- Are you interested in a comparison among instances to identify differences in implementation?

If so, continue on as QCA may be the right approach for your needs. If you determine that QCA is not the right method, there are some other evaluation methods to consider.

If the evaluator or implementer is looking to identify good practices and potentially use a more facilitative approach with stakeholders, but not rigorously analyze and compare information across instances, then we would recommend looking into the Appreciative Inquiry approach. Appreciative Inquiry is more improvised and amorphous in structure to allow implementers to explore ideas from a strengths-based lens. Comparatively, QCA is a more rigorous, structured approach to move beyond exploration into analysis to answer a set of specific questions.

Related to Appreciative Inquiry and QCA, is also the Positive Deviance (PD) approach, which is useful if an implementer wants to understand only the best possible outcome(s) of an intervention. Positive Deviance could be layered within QCA if the evaluation questions indicate a need or desire to move in this direction, but evaluators should be mindful that both methods accomplish different goals—QCA to compare factors among cases, and PD to identify and detail successful outliers.

In terms of use, if an implementer or evaluator wants to compare slightly different iterations of an intervention formatively to identify elements of implementation for scaling, we would direct readers towards quasi-experimental methods or other learning methods like Rapid Feedback Experiments (RF). The nuance with RF is that the threshold for action is "good enough" learning to rapidly change something iteratively, not as a summative approach to look retrospectively for answers.

Meanwhile, those with a need for controls and comparisons to prove formal hypotheses should proceed along the path of Randomized Control Trials (RCTs) as that method introduces additional components in experimentation for rigor that QCA does not include. RCTs are used to affirmatively prove or disprove a hypothesis or explanation as a result of variables, whereas QCA is looking to compare those variables or factors to a less rigorous degree for enhanced understanding.

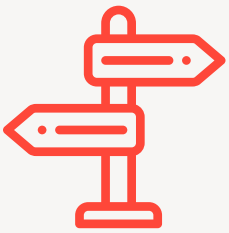
If an organization is seeking to understand and work out a theory of change for new, innovative, or complex activities, or understand what outcomes they have contributed to, then Outcome Harvesting (OH) is a more appropriate approach. Outcome Harvesting can be embedded within QCA in design, but each method unearths different types of information—OH yields intended and unintended outcomes, substantiation data, and contribution information, whereas QCA looks between identified factors to understand what may have influenced the end state, without fully validating outcome causal mechanisms in and of itself.

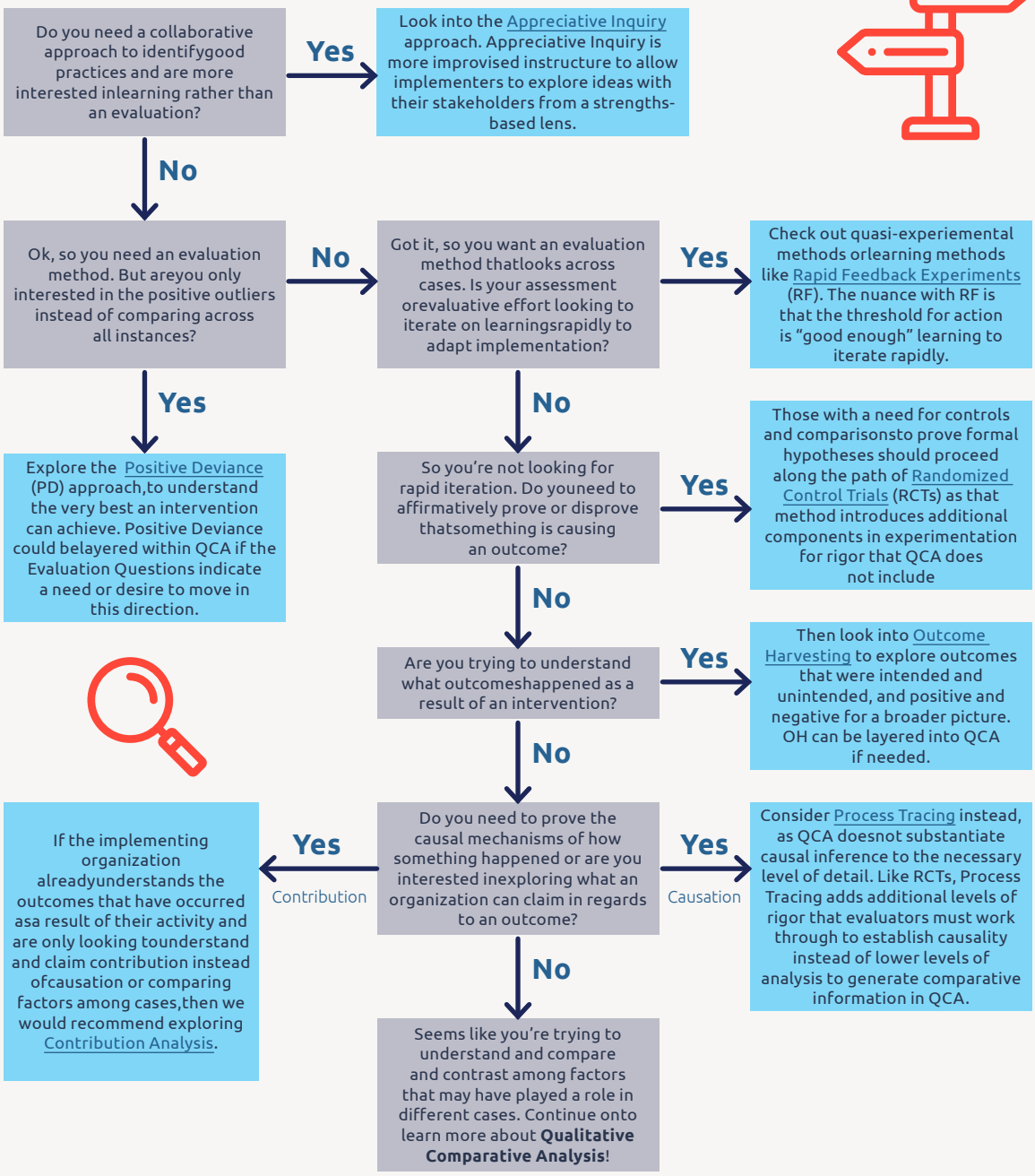
Finally, if the organization or evaluator is looking to rigorously understand and test the causal pathway of how a known outcome was achieved, Headlight would direct them to look into Process Tracing instead, as QCA does not substantiate causal inference to the necessary level of detail.[5] Like RCTs, Process Tracing adds additional levels of rigor that evaluators must work through to establish causality around both the change pathway and influencing factors instead of generating comparative information in QCA.

Headlight has provided a quick evaluation methods Decision Tree below focused on the implementer's questions that might lead to selecting QCA or other methods around similar learning and accountability aims. This Decision Tree is not exhaustive and anyone making this decision can read up on evaluation methods on BetterEvaluation.org, through various texts available on individual methods, or through Headlight's existing and forthcoming practical application Methods Memos.

5 Process Tracing is a qualitative analysis method that works to establish whether and how a potential cause or causes influenced a specified change or set of changes (Intrac, 2017).

Methods Decision Tree: Comparing and Contrasting





LIMITATIONS TO QCA

An evaluator must have a clear understanding of a method's limitations before finalizing their selection of an evaluation method.

One fundamental limitation of QCA is the comparability of cases – the cases must have a core component in common to qualify for QCA. The common component could be the type of intervention,

the desired outcome or objectives, the beneficiary population, etc. QCA is not just a compilation of case studies and pursuing this evaluation method with a random assortment of cases will prevent useful analysis and comparison of factors that could provide actionable findings.

QCA is also limited in the type of broader questions it is able to answer. QCA can uncover *what* changed in some cases and explore information about *why*, but the method does not detail the specific causal mechanisms of *how* the change occurred by itself.

Additionally, implemented on its own, QCA will not substantiate identified outcomes or results (although this can be achieved by embedding Outcome Harvesting as a core factor).

And lastly, QCA is not designed to validate a theory of change but findings from applying this method can increase understanding of the implementation of the theory (e.g., what did the change look like – but not if the change was achieved according to the theory of change).

MANAGING EXPECTATIONS

This section includes a brief overview of expectations around implementation and evaluation results to discuss with the client during design (more on implementing the method is in Module 3).

To begin, an evaluator/client should expect the budget and timeline for the evaluation to depend on how many cases are included and the number of factors. Each case requires individual data collection which has time and cost implications – the more cases, the longer the data collection and analysis phases and the more expensive in terms of labor cost for evaluators, potential travel to multiple locations for data collection, etc.

Another precursor to getting started is doing some type of evaluability assessment to determine the accessibility of reliable and credible information. Conventionally, evaluability assessments are done to ensure that the timing and environment are right for actors to undertake an evaluation. If so, then evaluators work to identify what the most suitable approach is to ensure that efforts can be contextualized. In the case of QCA, it is important to ensure that data sources will provide the necessary information on each selected factor and are accessible to the evaluator. An evaluator can assess if the needed information is sufficiently available by determining what data they will need to answer the evaluation questions and if/how they can access the data. If the data is not available or is inaccessible, this should influence factor and case selection.

At the end of a QCA evaluation, the implementer can expect to gain a more nuanced understanding of how different contexts affect either a particular intervention approach or a type of outcome. Analysis of the factors typically results in information such as, "the strength of XYZ coordination practices among donors influenced the intervention's success in six of the ten cases" or "positive political will, often realized through active engagement in and championing of intervention efforts by at least one local ministerial office is essential to efficient and effective implementation as seen in seven of ten cases". The client can use this information alongside implementing organizations to make pivots if the activity(s) is ongoing and/or inform future programming (e.g., invest more money in donor coordination working groups because that influenced success in a majority of cases, enhance early and often engagement with local ministerial offices as a key stakeholder, etc.).

MODULE 3:

Implementing QCA

Once an evaluator has determined that QCA meets their evaluative needs, it is time to begin fuller design[6] and implementation. Method selection should drive all aspects of implementing the evaluation – from design to analysis. The first part of this Module details how to design a QCA evaluation beginning with selecting the cases and factors, identifying data sources for the needed information, building a sampling strategy, drafting the interview protocol(s), and planning mitigation actions for any assumptions and limitations. With a clear design and plan in place, an evaluator can move to the data collection phase, and then on to coding and analysis. Each of these phases is described in detail in this Module with practical tools and guidance throughout.

DESIGNING THE EVALUATION

Case Selection

For some evaluations, the implementing organization will have cases in mind that they want to be evaluated. To stay within the scope of the method, ideally, 10-50 cases will be identified, and the evaluator may need to narrow down the list with the client depending on the budget and the desired or available timeline. Typically, when deciding that QCA is the best-fit method, a client/evaluator should already have a natural idea of the cases, maybe not the specific list but a general sense of what they should be to answer the evaluation questions.

Embedding Another Method

In some evaluations, during the method selection or the design phase, an evaluator may identify that QCA alone will not answer their evaluation questions and they need to embed another method into QCA. It depends on what the other method is, but generally, when answering evaluation questions requires more than one method, an evaluator should integrate the two to minimize the data collection burden and best leverage available data sources. When one of the two methods is QCA, it is appropriate to embed the other method within QCA because QCA requires a more expansive approach in comparing the selected factors across all chosen cases. The other method should be integrated as a factor so that it can be comparable across the cases. For example, if combining QCA and Outcome Harvesting (OH), an evaluator would fit OH, the more narrow method, within QCA, as a factor looking at what outcomes are generated from a similar intervention strategy across cases. When collecting data on outcome-related factors, QCA would then serve to identify the outcomes, and Outcome Harvesting to build out the outcome narratives and substantiate the outcome(s) occurrence, significance, and implementer's contribution to that outcome.

6 All design components should be captured in an inception report. An evaluation inception report serves as the grounding document for designing, planning, and implementing the evaluation. An inception report should include the background context of the evaluation, objectives and purpose, evaluation questions, detailed methods and procedures (overarching evaluation design, assumptions, limitations, sampling strategy, data management plans, description of analysis methods, etc.), expected key deliverables, an evaluation workplan, and data collection tools.

The most important consideration is the similarity of the cases to be compared – they need to be similar enough that comparing based on the factors will yield results that can inform future decision-making, pivots, etc. The conclusion cannot be that the cases are all different – otherwise, this method would just be a compilation of unrelated case studies, without the ability to draw meaningful conclusions.

Case selection also needs to account for triangulation across key factors. The key factors are those that you are trying to test and which have a definitive answer and therefore need some level of triangulation for the comparison to be valuable. Evaluators do not need to reach a statistical significance, but they will need a minimum of nine cases to properly assess one factor (i.e., three from Locale A, three from Locale B, and three from Locale C) to enable a minimum of triangulation within the locales and comparisons across the three groups. Base factor differences, like demographic characteristics, need to be mitigated where possible to enable a useful comparison against the key factors unless contextual differences are the main focus of the evaluation.

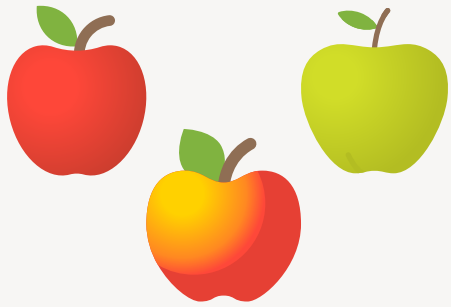
How to Take a Data-Driven Approach to Selection

The final set of cases selected can be driven by many considerations, and which takes precedence depends on the timeline for the evaluation, needs of the client/implementing organization, accessibility to potential interviewees, etc. For instance, maybe the implementing organization's goal is to compare as many cases as possible but there are budget or time limitations that require the evaluator to take a representative sample from the broader list.

If an evaluator has a broad list of cases but needs to narrow them down to align with the available resources, they should begin by conducting some quick analyses on the potential cases like running pivot tables to easily see the breakdown by characteristics of interest e.g., geographic location, amount of money in the activity, sectoral focus or intervention strategy, population size, etc. This step should be taken in close communication with the client/implementing organization to identify their priority areas of interest and discuss trade-offs of focusing on some over others. Choosing the cases by a representative sample is one way to narrow the list, or if there are cases the implementing organization is most interested in, the evaluator should ask questions to understand why that specific list and ensure there is a balance of cases likely to have both positive and negative findings.

Even if the list of cases from an implementing organization is pre-determined and of a manageable size, ensuring there is sufficient similarity between the cases to enable worthwhile Comparative Analysis is essential. In general, cases should have a core component in common that is closely related to the identified knowledge gap or desired information that led to conducting the evaluation. If the client is most interested in how contextual factors affect a particular type of outcome, an evaluator would want to ensure that the cases come from different geographic locations but have very similar intervention strategies or theories of change. By comparison, if the client is most interested in how different intervention designs influence achieving similar desired outcomes, an evaluator would want to ensure the implementing contexts had very similar variables and the implementing organizations had similar resources available for implementation. The figure below uses a comparison of different fruits to visualize how an evaluator should look for similarities in selecting comparable cases.

Selecting Comparable Cases



Example 1

Comparing different types of apples

If the core component in common is the type of fruit (apple), then the evaluator can compare a set of factors common to all apples but that differs across apple variety, e.g., number of seeds, color, average weight, etc. Evaluators cannot meaningfully compare factors when there is no commonality in cases (e.g., comparing the average weight among apples, oranges, and watermelons).

Example 2

Geographic comparison of where apples come from

If the apple variety is the same, but the geographic location varies, an evaluator can learn more about the contextual and environmental differences in the same apple variety in different parts of the world, e.g., flavor profile, type of soil grown in, average sales per year, etc. This comparison of a factor can help the evaluator assess which factors contribute to differences (i.e., a red apple from the same seed may taste different from Poland and China because of environmental factors).

Example 3

Different types of stone fruits

If the common component among cases is that they are all stone fruits, an evaluator can compare a variety of factors across the different fruits like the average size of the pit, region of origin, popularity in a certain country, etc.

Some questions the evaluator can ask to check the similarity of cases and assess how they will be compared:

?

- Which evaluation question is the highest priority to answer? Does the list of cases enable comparison and analysis to answer that question?
- How alike are the preliminary cases? What do the pivot tables show regarding the breakdown of characteristics the implementing organization or client is interested in?
- Where the cases are different – is that of interest to the evaluation? Do the differences align with early comparative factor selection, or are they so different that comparison across factors will not yield actionable results and analysis?

Case Selection Q&A

Q **Where does the recommendation that QCA includes ideally 10-50 cases come from? Is there any distinction between using QCA for research versus evaluation, acknowledging that a high number of cases is less feasible for most evaluation scopes?**

A This is the recommended practice from Charles Ragin's source material (Ragin, 1984), not necessarily the precise method application we are advocating for as it is not inclusive of other sampling saturation research (Guest, et al., 2006). Application of QCA with quantitative data collection makes it more feasible to do a greater number of cases; however, it risks losing significant comparative nuance as to why the differences or similarities between cases exist. Another contributing consideration is whether the study is being implemented for research or evaluation as research may not have the same time considerations or resource constraints that an evaluation does. As long as you ensure you will be able to triangulate findings as a bare minimum when determining case and factor selection, that is the guidepost, typically resulting in 12-16 cases for a typical evaluation of development programming.

Q **Is nine the minimum number of cases to compare? Can an evaluator compare fewer cases?**

A Thinking about sampling saturation standards, Headlight would advise implementers to lean on standards of triangulation. Per case type, the evaluator would need a bare minimum of three cases to identify triangulated trends within the case type. Ideally, the evaluation would have more than three per case type because it is unlikely that all cases are identical or have sufficient data on all comparative factors, so more cases give space for more trends to appear within a case type. Because QCA is comparative, there need to be multiple case types to look across factors/variables. The absolute bare minimum would be six cases, three cases of one type, and three of another, but the chances of triangulating findings with these few cases is low.

Factor Selection

After an evaluator has confirmed case selection, it is time to choose factors for comparison. Factors are aspects, characteristics, or features that vary among the cases and which the evaluator uses to conduct cross-case comparisons. The types of factors needed for a QCA evaluation can be broadly split into two categories – 1) demographic or environmental characteristics which can be contextualized in many ways but need to be included so you can accurately compare the cases and 2) factors that identify what you want to explicitly compare (results, approaches, etc.). Sometimes these factor sets can overlap, but not always, and an evaluator needs to ensure a balance between them.

The table below provides examples of different factors, by factor categories, and is meant to provide illustrative examples only.

Examples of QCA Factors by Factor Categories				
Location-Based	Time-Oriented	Process & Implementation	Outcomes	Influencing Environment
Level of implementation (national, regional, local) Proximity to the nearest health center Population size of the intervention area	Length of intervention (period of performance) Timeliness of intervention (in response to the issue/challenge it seeks to address)	Adherence to the proposed intervention scope Coordination with other stakeholders Use of monitoring, evaluation, and learning systems	Intended and unintended outcomes 7 Change process leading to set outcomes	Level of local community engagement Cultural and social norms External shocks or crises

The evaluator should start by considering the evaluation questions. Per question, an evaluator should ask:

?
- What factors and/or categories of factors will help answer the question?
- What do we need to understand about each case to answer the questions?
- What information will provide enough context and nuance for those answers to be actionable?

For example, if the selected cases represent different intervention sites and the question is "What were the greatest differences in factors that influenced X intervention in XYZ intervention sites? How did those differences affect the implementation timeline?", an evaluator should begin by identifying the location or influencing environment factors across the intervention sites. This will depend on the overall context and nature of the activity/intervention, but factors of interest may include: the population size of the intervention site, proximity to the nearest city, level of community engagement, historical implications on social norms, and even topographical differences when focusing on more agriculture or climate-based interventions, etc. An evaluator would also want to include a factor related to the timing or length of the intervention/phases of the intervention in alignment with the second half of the example evaluation question.

When identifying factors, there are two common pitfalls.

1. The factors are too broad and only or mostly related to environmental factors. QCA evaluations are meant to focus on a specific set of cases and factors, not to capture the entire system. The evaluator needs to keep that in mind when designing or risk not being able to capture data on the factors or make use-focused comparisons across the cases.
2. The evaluator leaves out factors related to the results and supporting components to understand the results, which enables the implementing organization to take action. There needs to be some orientation of the factors on results and the "so what" for the implementing organization. Much of the "so what" will come out in the analysis phase, but results-related factors will ensure this data is captured to be analyzed later.

7 Understanding how different outcomes evolved, how actors contributed to them, and whether or not they are substantiated requires a separate or embedded implementation of the Outcome Harvesting methodology. See more about Embedding Another Method on Page 17.

Traditional Approach v. Headlight's Approach to Comparing Factors

The traditional approach to comparing factors in QCA is to use a numerical assignment like a 0 if the factor is not present and 1 if it is present in the case, rank each factor on a Likert scale, or choose from a predetermined list of options for each factor. Headlight's variation is not as prescriptive and enables the collection of information that does not fit a numerical value, such as qualitative factors like "coordination among stakeholders." A factor like that could be analyzed by assigning 0 if stakeholders do not coordinate, and 1 if they do. However, there is a lot of information that would be missing in that binary analysis. This approach is more in line with the value of the qualitative aspect of QCA– preventing a false quantification of qualitative information.

Headlight's approach to a factor like "coordination among stakeholders" would be to use Thematic Inductive Analysis (more information in the Analysis Section) to identify trends, barriers, bright spots, and sub-themes to what coordination does or does not look like among stakeholders in each case. What was the most common type of coordination mechanism in each case? What was an example where a lack of coordination had a negative impact on the intervention's success? What conditions enabled productive coordination? Applying additional qualitative analysis methods when comparing the factors enables the evaluator to more fully assess and understand the "why," "how," and enabling and inhibiting environments to the cases, as well as collect data that stakeholders can use to implement adaptations based on the evidence.

TIP: It is useful to set up the evaluation questions in a whiteboard space (e.g., MURAL, Miro, or in-person whiteboard) and then brainstorm factors related to each evaluation question. This helps to visualize the connection between evaluation questions and factors. Annex A presents an example template for how to set this up.

Once an evaluator has brainstormed the relevant factors, they should go through and determine which identified factors are necessary to answer the evaluation question versus which may be merely nice to know, which are specific enough for comparison, and if there are enough factors to answer the evaluation questions. Several considerations to keep in mind during this step are:

- The number of factors selected is associated with at least one question in the interview protocol if not more – so ten factors are at least ten interview questions, likely more with follow-ups and additional prompting questions. Make sure the total number of factors multiplied by the total number of cases and anticipated data collection burden aligns with the available resources and a reasonable ask of time for interviewees.
- Depending on how many factors have been identified, it may be useful to sort them by category. This can be derived from the evaluation questions and anticipated areas for analysis. For example, the question above about intervention site differences may inform a category of factors related to "characteristics of the intervention sites." This can help sort factors into groups that will require similar data collection efforts and/or analysis methods.
- During the selection and refinement of factors, consider which factors enable the best comparison across cases. It is important that factors will lead to information that tests a larger hypothesis or theory of change and reveal more information than just the conclusion that each case is unique.

Data Sources

Before the factor selection is final, an evaluator needs to identify the data sources for information on each factor, e.g., secondary documents, key informant interviews, etc. This process can also be done in the whiteboard space. Once the factors are mostly set, an evaluator should look at each one and assess where they have sufficient access to the necessary data to enable comparison, often in consultation with the client/implementing organization. As shown in the template in Annex A, it is useful to also denote in some way what data the evaluator already has versus what they need to collect.

Sometimes after determining and assessing data sources, the evaluator may need to revisit and further refine the factors to ensure the feasibility of collecting useful data for the evaluative effort. For instance, if it becomes clear accurate data may not be available across cases for a particular factor or there are essential people that they do not expect to be available or accessible to speak with it may be necessary to drop those factors and manage expectations around anticipated answers to associated evaluation questions. For example, Headlight's Strategic Advisory Group for Emergencies (SAGE)/Crisis Modifiers Evaluation for USAID/Ethiopia had identified several time-oriented factors that required secondary documents as the data sources, and though the evaluation team thought these would be easily accessible during the design phase, the data did not exist for all the cases, which meant that the factors that relied on those data sources could not be fully analyzed and time-oriented evaluation questions could not be answered.

Sampling

Building the key informant interview or focus group discussion sampling strategy is dependent on the case and factor selection – who will have information on these, what parts of the cases are best known by whom, where there will be differing experiences and perspectives, etc. Headlight recommends a combination of purposive and snowball sampling to attain a rigorous and use-focused sample for a QCA evaluation. Purposive sampling should be used for the initial data collection, based on which interviewees are best informed about the cases in the evaluation. Snowball sampling will then be used to achieve saturation (explained below), relying on planned/existing interviewees to identify additional people to speak with, keeping in mind the diversity of stakeholders represented.

The evaluator should leverage an early desk review during the inception report drafting phase, alongside ongoing conversations with the client/implementing organization, to identify who all the essential stakeholders are, which category they might fall into, and how to maximize collecting information from different perspectives within the data collection limitations and available resources. The interviewees are likely to fall into three categories:

- **Definitive interviewees:** this group includes people that the evaluator assumes to have **detailed and specific knowledge** of the cases.
- **Limited access and oversight:** this refers to a subset of a broader group that the evaluator believes to have **some relevant knowledge** pertaining to the evaluation question(s) and cases. Not everyone in this group will have specific or complete knowledge of the cases. This group is likely to be larger than the definitive interviewee's group, but increasing the number of people in this limited access and oversight group does not necessarily increase the perspectives or quality of information received on the cases.

- **Homogeneous:** this refers to a group that likely has **a similar experience with the cases**. This group of stakeholders requires sampling saturation and is likely to overlap with the two categories highlighted above. Sampling saturation can be achieved by conducting a minimum of six interviews or three focus groups per homogenous group, which will achieve approximately 70% coverage of unique information and ideas (Guest, et al., 2006). This number is appropriate based on the sampling saturation standard of between 6-12 interviews to reach 70-92% coverage of unique information and also accounts for the need for timely and use-focused data collection. If going beyond purposive sampling, evaluations are likely to have several homogenous groups that need to be accessed during data collection and all such groups should reach sampling saturation minimums in qualitative efforts.

Example of a QCA Sampling Frame

Going back to the example of apples in three different countries that grow the same type of red, classic apple, the following table presents an example of a sampling frame that evaluators would use to determine the scale of data collection efforts across stakeholder groups. The sampling frame below is built as an example and is not designed to be prescriptive or a standard sampling frame for any QCA effort. Stakeholder groups should always be chosen in alignment with evaluation questions and other design considerations.

Example Sampling Frame Table

Cases [8]	Homogeneous Group 1	Homogeneous Group 2	Definitive Interviewees	Limited Access Interviewees	Total Interviewees
Farm A in China	Farmworkers – 6 people w	Customers – 6 people	Seed suppliers – 2 people	Grocery store owners – 4 people	18 people
Farm B in China	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 3 people	Grocery store owners – 4 people	19 people
Farm C in China	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 2 people	Grocery store owners – 5 people	19 people
Farm D in China	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 3 people	Grocery store owners – 4 people	19 people
Farm A in Poland	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 2 people	Grocery store owners – 4 people	18 people
Farm B in Poland	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 4 people	Grocery store owners – 2 people	18 people
Farm C in Poland	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 1 person	Grocery store owners – 4 people	17 people
Farm D in Poland	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 2 people	Grocery store owners – 3 people	17 people
Farm A in the United States	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 3 people	Grocery store owners – 4 people	19 people
Farm B in the United States	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 2 people	Grocery store owners – 3 people	17 people
Farm C in the United States	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 4 people	Grocery store owners – 2 people	18 people
Farm D in the United States	Farmworkers – 6 people	Customers – 6 people	Seed suppliers – 2 people	Grocery store owners – 4 people	18 people
Totals:	**72 people**	**72 people**	**30 people**	**43 people**	**217 people**

8 Farms would have to be selected within a more narrow regional band with limited contextual differences and high contextual comparability given the size and environmental differences within the example countries listed.

Starting with the cases in this example, four farms were chosen from each of the three countries because, typically in QCA, at least one case per location or country may not have complete information anticipated or will not be as ideal of a fit as the evaluator assumed during evaluation design. Selecting four cases per country gives the evaluator a minimum buffer to try and ensure that they will be able to triangulate findings within each country so that they can more accurately compare between the countries as well.

Next, for the homogeneous groups of farmworkers and customers, these are both large pools of people to draw from for the sample, so the evaluator will leverage sampling saturation research (see the Sampling Section) and pursue at least six interviews for each of these stakeholder groups. As cited in Guest, et al., 2006, six interviews is the acceptable minimum for approximately 70% saturation in original concepts or ideas that are shared in response to the interview questions. Given the size of these groups, a random sample could be used for interviewee identification.

For the definitive interviewees, full saturation is achieved at the number given for qualitative interviews because these are the total number of seed suppliers supplying seeds to each farm, for the red, classic apple. The evaluator cannot use the aforementioned sampling saturation rates, because this is a definitive group (i.e., for Farm A in Poland there are only two seed suppliers that supply red, classic apple seeds for that farm). Because the group is definitive and small, the evaluator is choosing to speak with all the definitive interviewees, which will enhance their ability to triangulate when possible, and if not, to have more reliable inputs despite a low sample size.

Lastly, for the limited access interviewee groups, sometimes the evaluator may not have sufficient connections, social capital, attention, or interest from all the stakeholders involved in the topic they are evaluating. When that is the case, the evaluator needs to be forthcoming with that limitation both in how they plan for the sampling and how they will address the limitation in analysis. For example, in this apple situation, the evaluator could note in their inception report that the grocery store owners are a few steps removed from the farmworkers, who are helping the evaluator make connections with the other stakeholder groups, so they do not plan to pursue sampling saturation for this group due to the challenges with making connections. However, the evaluator plans to speak to as many grocery store owners as possible to best enable triangulation of findings.

Assumptions, Limitations, and Mitigating Actions

QCA enables strong Comparative Analysis to be conducted for a number of factors across cases, but the analysis potential of the method is underpinned by a number of assumptions:

Assumption: The evaluator will be able to capture sufficient documentation from the client(s) to support enough background knowledge to design informed evaluation questions and corresponding data collection protocols. Similarly, those involved in the activity or effort's implementation will agree to meet with the evaluator team and share any relevant information that supports their understanding of the how and why of implementation, as well as contextual factors.

Mitigating Action: If clients or respondents are unwilling to share information about a topic, that will understandably limit the information evaluators are able to work with and analyze. Evaluators should think about these dependencies while choosing, and potentially reselecting, factors and designing the interview instrument to ideate where else they may be able to source this information. Additionally, this dependency should be made clear to the client ahead of time to manage expectations accordingly.

Assumption: The evaluator should be able to identify enough qualitative data participants (interviewees or focus group discussants), with the client or implementing organization's help as warranted, to ensure **sampling saturation for selected homogenous groups** is met.

Mitigating Action: Evaluators may not always be able to achieve saturation for all groups in the event that there are not at least six individuals who would know about the topic being evaluated, assuming that the sample is from a large population. If this is the case, they should note this limitation in any final reports for sub-groups accordingly. Noting this limitation allows readers to acknowledge the strength of the evidence and situate any findings appropriately. However, if the sample is from a small population, let's say only four specific people were directly involved/provided X service, and you speak to each of those four people, then your evaluation is not less rigorous because you spoke to all people in the population (100% saturation). One key thing to keep in mind when choosing methods and determining sampling strategies is the importance of speaking with **enough stakeholders from a variety of groups so that the understanding, nuance, and representation in the findings can speak to how all the core stakeholders experienced the interventions' benefits or possible harms**.

Assumption: The cases selected for QCA evaluation are truly comparable.

Mitigating Action: While there is little to do to mitigate this once cases and factors have been chosen, evaluators should identify why a particular case is not comparable to others and reconsider cases included to ensure that findings will be actionable. Not all cases are a good fit, even if a particular example resonates with the client. Some cases are not able to be compared. The evaluator may catch this at the case selection phase and then can work with the client to choose another case. However, if not caught until analysis, then the evaluator may be only comparing a subset of the cases and need to leave out a particular case or factors with insufficient data and/or inaccurate comparisons. See the Case Selection subsection above for more.

Often in QCA, interviewees may have a perception, positive confirmation, and/or recall bias about the chosen cases in the evaluation.

With perception bias, interviewees allow their perception of the cases, factors, or context to influence their responses.

Mitigating Action: While this is not fully avoidable, evaluators should take care to design the protocol/interview questions in such a way as to mitigate bias and build triangulation and substantiation into their analysis plan. One way to do this is by ensuring questions asked are neutrally written, and case comparison is left to analysis versus hinted at with qualitative data participants.

Positive confirmation bias can happen if the respondents assume the cases were chosen because they are successful examples, or that the evaluator is only seeking positive information about a case. This skews the overall data, especially in that case.

Mitigating Action: The enumerator should explain in the initial email as well as reiterate during the interview or focus group discussion what selection criteria were used to choose the cases, if applicable. For example, the enumerator could explain that based on the evaluation objectives, cases were chosen from a variety of geographic areas, with key differing characteristics. Explaining a bit more about the selection process will allow the

respondent to see that a variety of cases were included, and the focus of the information gathering is on key features or characteristics, not trying to identify which case is the 'best'. Another way to mitigate positive confirmation bias, is again to ensure careful, neutral wording of questions, appropriate time to ask about challenges, avoid starting interviews or focus group discussions by asking about results immediately, and be careful not to mention the opinions of others regarding the case.

Recall bias may happen if some or all of the cases selected have interventions that concluded several years before the evaluation effort, and respondents struggle to recall the level of detail required for QCA.

Mitigating Action: The most important mitigating action to address recall bias is to choose cases that are more recent (ideally within a year since the conclusion of the intervention). If that is not possible due to the selection criteria, the enumerator should prepare a brief, neutral summary of the case and provide that in a written email as well as describe the case at the beginning of the interview or focus group discussion to help refresh the respondents' memories. However, the enumerator must be careful not to sway the respondents in any particular direction or include opinions or leading statements in the summary. For example, if the cases are different activities implemented by a donor, do not say, "this activity was one of the most challenging to the donor because they lost their funding…" but rather, neutrally state the facts that are needed to remind the respondents of the case. "This case involved XYZ interventions, implemented by this actor, during this time period, in these areas."

Data Collection Tools and Protocol

Once evaluators have gotten a sense of data sources, identified their sampling strategy, and acknowledged the assumptions and limitations, it is time to build the protocol and associated tools for use in data collection.

First, the evaluator develops a full data collection protocol. The protocol is the overarching guidance document for the data collection effort and should include subsections on General Facilitation Notes and Guidance, Informed Consent Language, the full list of Interview Questions, potential prompting questions to dig deeper, the desired template for note-taking, and a Wrap-Up blurb to share with interviewees about what will come next and what, if anything, they should expect regarding a report out. The protocol should also include different variations on questions, framing, or additional context needed for different stakeholder groups that will be involved in the data collection. Having the broader protocol document in place helps ensure consistency, especially when data collection is being implemented by more than one enumerator.

When building the data collection tool for a QCA effort within a protocol, Headlight most often uses the semi-structured interview approach where evaluators leverage a guiding questionnaire to facilitate a conversation around core themes and questions. But, focus group discussions are also an important data collection method to consider. For both semi-structured interviews and focus group discussions, questioning through the tool may veer slightly from the list in the instrument in order to ask follow-up questions to probe for additional details. When developing the instrument, there should be at least one question per QCA factor selected that can't be verified through secondary data sources. As such, focusing on need-to-know information versus want-to-know information becomes increasingly important. Otherwise, the instrument will become too lengthy, which lessens the likelihood that the evaluator will be able to get the respondent(s) to answer all the questions

in the available time, as well as places a larger burden on the respondents than may be necessary to receive for actionable, useful data. Some questions to explore when prioritizing data collection questions include:

?
- What information is necessary to answer the Evaluation Questions?
- What information is nice to know about but can be captured another way or at another time?
- What other sources of information might the evaluator have access to that complement, minimize, or reduce information needs from interviews in order to answer this question?

Prioritization and cross-checking that the questions cover the need-to-have information to answer the original evaluation scope are crucial, otherwise, data collectors may misalign their tools and come back with information that is interesting, but not necessarily needed by the client.

While the instrument may be aligned with the overarching Evaluation Questions, it is important for enumerators and data collectors to check in regularly during the collection phase to ensure that the responses participants are providing are sufficient to meet the evaluation's needs. Some of this can be avoided early on by including more detail in the protocol about what kinds of responses the team is looking for, suggested follow-up questions, and practice amongst enumerators to ensure the questions are interpreted and understood as intended. For example, perhaps a question is "How soon after the award was approved did the activity begin to respond and intervene?" If the enumerator does not know they should be looking for an estimated number of days as a response from the respondent, they may allow a respondent to share "very quickly" and not probe further for more specific information. Ensuring questions are specific, contextually accurate, and appropriate, and that enumerators are well-informed about the level of nuance to be captured is critical to the success of the evaluation. Regular check-ins with the evaluation team and quality assurance review practices can help evaluators avoid heartache and work that needs to be redone later on. While this is not a challenge specific to implementing QCA, it does hold more weight as comparative analyses require similar formats of information to enable the analysis to proceed.

DATA COLLECTION

Data collection can start as soon as data collection tools have been completed, sampling frame key informants have been identified, enumerators have been vetted and trained, and any additional resources and logistics have been appropriately procured/arranged. When collecting data, evaluators should be continually mindful of any potential limitations identified during the tool development stage. Evaluators should ensure that they are getting useful information and responses through periodic check-ins and quality assurance/quality control checks. This helps to ensure that responses match what was intended for in the initial design to answer evaluation questions, or in the event that responses are not as evaluators had hoped, this gives the team the opportunity to make any changes to the tool moving forward to improve quality data.

Selected interviewees may have information about multiple cases depending on case selection, which can pose complications around time management and recall bias. Evaluators can schedule additional time (e.g., 90 minutes instead of 60 minutes) and tailor the interview questionnaire to include follow-up prompts based on the specific case in question. Interviewees may not have a point of delineation if the cases in question overlapped at any point, and by reminding interviewees of details (e.g., time, context details, the name of the cases), then interviewees might be able to sort

through their memories more effectively. In the event that this is not possible, it is important for the evaluator to make note of this while transcribing interview responses so that this can be revisited in the analysis phase.

Lastly, data collection at scale for proper sampling saturation can be challenging, so evaluators should consider using a tracking system to monitor who they have sent invitations to, who they have followed up with, who they have spoken with, and any substitutions or adaptations made along the way for proper documentation and as a general best practice. Having a regularly updated tracker can also enable evaluators and team members to divide interviewing responsibilities more easily and also start coding on a rolling basis instead of waiting until all data has been collected.

CODING

This subsection specifically focuses on the nuances of qualitative coding for QCA.[9] Even when the factors have quantitative data to be analyzed, qualitative coding helps to illuminate the specific nuance of why something did or did not work, what factors influenced the timeliness of the intervention(s), etc. The specifics of the coding process will look different depending on the factors, which can influence the type of software that is best suited for the effort, however, software choices can also be dependent on preference and organizational licenses.[10] For the purpose of orienting the coding guidance herein, Headlight uses Dedoose for qualitative coding.

Once the evaluator has selected the qualitative analysis software they will use, the next step is to develop a codebook. Similar to steps in the evaluation design, the process of developing a codebook begins with the evaluation questions and the factors. Designing codes based on the factors will ensure not only that the evaluator is coding for the right types of information that will answer the evaluation questions, but also make clear which codes to analyze during secondary analysis for each factor. For example, for the factor "use of monitoring, evaluation, and learning (MEL) systems," the evaluator may have identified codes like "What Works /MEL systems" and "What Doesn't Work/MEL systems." Using these codes, the evaluator would be able to identify examples of MEL systems that work, MEL system challenges, and how they know the systems work across data collected for all cases.

The evaluator should be aware that codes are likely to also emerge throughout the coding process. Emergent codes are typical for most qualitative coding efforts, but will be especially important if the evaluator is embedding another method within QCA like Outcome Harvesting. Creating mental space for emergent codes ensures that sufficient nuanced details and unexpected trends can be properly identified and added during the coding phase without being too prescriptive at the outset and thus missing information that might be pivotal to understanding the cases.

Another key aspect of coding in QCA evaluations is to have a clear set of descriptors or tags for each document that the evaluator will be coding. Assigning these descriptors will allow the evaluator to identify exactly which case the data is associated with, which will assist in keeping things organized. Descriptors should be aligned to the case selection criteria and any other demographic-orienting information necessary to categorize and sort the data effectively.

9 For more information on some of the basics of coding, Headlight has two blog posts that share the top things to do and to avoid in qualitative coding, one from September 2020 and a follow-on piece from March 2022.

10 This blog post provides a detailed breakdown and comparison of several popular qualitative analysis software platforms, to assist evaluators in deciding what software is best for their evaluative and analytical needs.

The last and most important consideration for qualitative coding, which is true in any evaluation, but especially QCA given the structure of the factors, is that the evaluator is not coding for just anything that is interesting in the case but **specifically to evidence on the selected factors**. When designing the codebook, evaluators should confirm that they are identifying what are the "need to know" pieces of information and then during coding, that they are specifically coding to excerpts that provide the answers. Some internal prompts that an evaluator or coder can use to ensure this step is happening are to identify a quality assurance/quality control process at the outset of coding, to host check-ins with other coders along the way to discuss questions or difficulties, and to continue to revisit the codebook and reflecting on questions like:

?

- How does the coder know this works/doesn't work?
- Is what the coder is coding clear in its explanation of how a factor affects something?
- Does the excerpt speak specifically to one of the evaluation questions?
- How will the clients and audience use the information that the coder is coding?

ANALYSIS

Once coding is complete from all of the documents and interview transcripts, evaluators can then move into the analysis phase to find more nuance in what has been coded to answer their evaluation questions. At the outset of the effort, evaluators should have generated an analysis plan section in their inception report to strategize and document their approach to how they will use the information being collected. This can also be done during data collection once the evaluator sees the types of data, where there may be gaps, what information is hard to find, etc. so that they know best how to proceed.

Specific to QCA, the analysis plan needs to map each factor to an analysis method to ensure that the information is analyzed and used properly to answer the evaluation questions and compare the factors across the cases. The evaluator must consider which analysis methods are best suited to provide information on the factors. It is worth noting that inherent in the evaluation method of QCA is Comparative Analysis – all factors will be analyzed using Comparative Analysis methods but sometimes Comparative Analysis is dependent on other types of analysis as the foundation for more complex comparisons. The table below provides a few illustrative examples for selecting analysis methods based on the factors.

Mapping Factors to Analysis Methods		
Factor	Analysis Method	Rationale
Level of local community engagement	Thematic Analysis; Comparative Analysis	This is a complex factor that is likely to have many sub-themes emerge during analysis. The evaluator would use **Thematic Analysis to look for nuanced information** about the level of engagement, how that was achieved, enablers/ inhibitors, etc., and then apply a **Comparative Analysis lens to how the level of engagement varied across cases.** The foundation of Thematic Analysis for this factor will allow the evaluator to also uncover ***why*** the cases were different and identify recommendations for what the client or implementing organization may do to respond to the findings.
Adherence to the proposed intervention scope	Process Analysis; Comparative Analysis	Process Analysis is a good fit to analyze this factor because the evaluator wants to identify which interventions adhered to the proposed scope(s) which outlines the necessary steps and procedures for implementing the intervention. Process Analysis will illuminate if there are steps that all cases followed, if any steps were pivotal to the success or failure of the intervention, and where implementing organizations can improve the process. Comparative Analysis should also be applied after the Process Analysis is complete to compare how processes were followed across and between the cases.
Timeliness of intervention (in response to the issue/challenge it seeks to address)	Comparative Analysis	To analyze this factor, an evaluator should use **Comparative Analysis to understand the difference in timeliness** of response across the cases. The more variance this factor has in the Comparative Analysis, the more likely that the evaluator would need to apply an additional analysis method to understand why the time is so varied across cases.

In the first type, Thematic Analysis, evaluators take the details of what has been collected in primary coding and conduct an additional round of Thematic Inductive Analysis on coded qualitative excerpts to find any additional layers of nuance. While different practitioners may have different practices, this section details the Headlight process. First, export any codes that the evaluator wants to conduct deeper analysis on into a Google Workbook. From there, emergently code for trends that are seen in one larger code. For example, maybe there are 50 excerpts that got coded to What Doesn't Work>Barriers and Challenges>Insufficient Resources. During secondary analysis, look through just these 50 excerpts and apply emergent codes like "Insufficient Human Resources/LOE" or "Lack of Access to Supplies," based on the themes seen within this excerpt subset. Doing this allows the evaluator to pinpoint what resources are insufficient and anything else about how this affects implementation, to inform targeted recommendations to the client. Once more nuanced thematic findings have been detailed for each individual case, then Comparative Analysis can be used to look at trends across and between cases.

In Process Analysis, evaluators start by exporting all the codes relevant to the process into a Google Sheet or Excel workbook. The first step is to sort the excerpts into broad categories of steps, i.e., pre-implementation, implementation, closeout and reporting, etc. Then within each of those categories, the evaluator begins to further sort the excerpt into distinct, sequential steps. Each step should be at a minimum triangulated across cases in a common group to confirm that this is a verified step. For some steps, it may be triangulated by individual case sources, instead of across similar cases, which would make that step an outlier for that particular case's implementation. This might still be interesting to include especially if there is additional Thematic Analysis on whether this case worked especially well or not so well. For all the triangulated steps, the evaluator should write up a short summary of the step, the data providing evidence for the step (i.e., how many sources confirmed this is a step, how many cases, etc.). Once the steps are clarified for each case, then a larger comparison can be drawn on outliers, similarities, differences, adherence or lack thereof to a process, etc. Similar to the other analysis types, the evaluator may find it useful to draft an FCR specific to the Process Analysis, since the steps function as the findings for this type of analysis, and the conclusions and recommendations may be specific to certain steps or contain information on the process as a whole.

When doing Comparative Analysis, evaluators start with the information that they are trying to compare. This might look like quantitative data (e.g., number of days for implementation by case) or qualitative data (What Doesn't Work>Barriers and Challenges>Insufficient Resources by case). Comparative Analysis requires determining factually what the differences are and then using additional findings and nuance from the data to try and explain why those differences and/or similarities exist (e.g., "The number of days from activity initiation to implementation ranged from 8 to 45. Organizations with quicker initiation times were more likely to be between 5 and 10 years old and to have prior experience with the donor"). While evaluators are unlikely to know the causation fully, they should capture some contextual understanding and what might be contributing to differences. More importantly though, evaluators should include something about the broader picture—e.g., "*faster initiation times were linked with XYZ trend, so the faster the response, the better for emergency response.*" Finally, evaluators should end with some recommendations like, "*with such a wide variation in initiation times, XYZ party should further explore ABC factor and ensure that they are not inhibiting RST part of the process. A particularly notable case to look into is JKL case as it had the fastest response time and also reported the fewest challenges at RST phase.*"

The type of analysis conducted for each evaluative effort should be dependent on the factors selected during evaluation design to help synthesize actionable information against the evaluation questions as each analysis type helps to answer a different angle of the questions (the analysis examples above are only illustrative). Once each type of analysis has been completed, evaluators should look at the FCR[11] as a whole to translate findings, conclusions, and recommendations into a use-focused deliverable.

11 For more information on some of the basics of writing a Findings, Conclusions, and Recommendations Matrix please see Headlight's blog from September 2020

Analysis Q&A

Q **Are there instances where Comparative Analysis is used in other methods? Does implementing Comparative Analysis automatically imply an evaluator is using QCA?**

A Analysis methods and evaluation methods are distinct components of how you structure an evaluation. The evaluation's overall structure in how it will approach answering the evaluation questions, how the evaluation intends to source data, and analyzing data for findings are all unique parts of the process. There are some analysis methods that are more likely to be used in conjunction with particular evaluation methods, but analysis methods are not associated with a singular evaluation method. For example, for Outcome Harvesting, you will almost always use Narrative Analysis because the end goal is to have outcome descriptions that require a sensical narrative of how the outcome happened, but Narrative Analysis is also used in Most Significant Change, Positive Deviance, and other evaluations. For QCA, you need to compare factors across cases, so at some point you will need to use Comparative Analysis, but it may not be the only analysis approach needed. Separately, you could use Comparative Analysis in a Realist Evaluation when assessing divergence from the hypothesized theory of change, in a Most Significant Change evaluation comparing the similarities and differences of prioritized changes by different populations, or in a multitude of different evaluation approaches. Just because an evaluator is doing Comparative Analysis, does not de facto make the evaluation a QCA evaluation. QCA is a structure for an entire evaluation, not merely how you respond to one evaluation question or assess data collected as part of an evaluation.

MODULE 4:

Evaluator Competencies for QCA

Competency Needed	Rationale	How to Assess
Skill: Knowledge of Qualitative Methods	Skills should include experience working on qualitative data and methods, understanding of how to ensure rigor when applying qualitative approaches, and direct prior application of QCA is obviously preferable.	Ask what qualitative methods an evaluator has used, looking for qualitative evaluation, not data collection methods. This will be an indication of their qualitative experience Write a draft qualitative approach as an exercise or have the evaluator explain a qualitative method they've used in an interview
Skill: Comparative Analysis	Experience working in comparative studies. Comparative Analysis skills can be from quantitative methods, but would need to be supplemented by qualitative experience (above).	Ask a situational interview question where the candidate needs to comparatively analyze information. Those procuring a QCA approach should look for answers with clear identification of comparative variables and how they would go about the comparisons (e.g., designing the comparison), not the evaluator demonstrating how to present findings.
Skill: Experience Contextualizing and Adapting Methods	If the original QCA assumption is 10-50 cases, the evaluator needs to know how to proceed in terms of the number of cases and the content of cases to meet the desires of the evaluation's scope. This ties into the two skills above in that an evaluator needs to be able to think through how to apply the method for appropriate case selection.	In a technical exercise, the client should provide a paragraph of an evaluation scope, a paragraph of context, and ask for a few brief reflections on how the evaluator would need to adapt and contextualize the method to meet the stated needs.
Skill: Facilitating Semi-Structured Interviews	Being able to conduct a semi-structured interview will enable nuanced data collection on a variety of factors with sufficient skills to follow up for more detailed responses as needed	Ask an interview question that gets to an evaluator's experience conducting semi-structured interviews (e.g., tell me about a time when you conducted data collection interviews. What did you learn? OR what do you do in an interview when the respondent doesn't provide enough detail to the question you have asked? How would you proceed?)
Topical Knowledge: Analysis Methods	Aside from Comparative Analysis, some QCA implementation efforts require multiple types of analyses in order to be able to build on one another. Evaluators should have a breadth of experience with different analysis methods (e.g., Thematic Inductive, Thematic Deductive, Process, Narrative, Comparative, etc.)	In a technical exercise, provide the evaluator with a table of QCA factors and ask them how they would go about doing analysis on a given factor. Evaluators can self-select/indicate which approaches they have experience doing. (**This is tricky to verify as it is a self-select) Ask an interview question that gets to an evaluator's experience doing different types of analysis (e.g., Tell me about a time when you analyzed data to answer a time-related or sequence-related evaluation question. How did you go about the analysis?)
Skill: Utilization-Focused	An evaluator should be utilization-focused throughout the whole evaluation from case selection to data analysis.	In a technical exercise, prompt the evaluator to identify stated needs or uses of the information based on the context provided.

MODULE 5:
Top Tips for QCA

This Module provides quick tips and reminders for implementing QCA. Although these tips are covered in the fuller design and implementation guidance of Module 3, this Module offers a refresher on common pitfalls and challenges for quicker reference.

1. CASE SELECTION IS CRUCIAL!

Typically when you begin evaluation design, the implementing organization will have an idea of what cases they would like to include in the evaluation, or at least a few suggestions for how to pick these cases. Preliminary case ideation is an excellent start and can help further refine evaluation objectives and questions, but the initial case ideas should not be agreed to without assessment to ensure they are **truly comparable and that they are within a reasonable range or number of cases** (more information in the Case Selection section). To do some quick sense-checking on both of those aspects, the evaluator should consider that the total number of cases is within 10-50, and depending on the budget and timeline, making sure that the number is right-sized according to those limitations.

During this scoping and refinement, it is imperative to understand and emphasize that adding one more case is not a low effort. Rather, in QCA, **adding an additional case has a multiplier effect**. Referring back to the red apple example, each case has approximately 20 associated interviews, so there would be a direct multiplying factor when adding even just one more case, and that also comes with additional time and budget constraints.

Additionally, the evaluator can **do some preliminary desk research and create pivot tables to see the different cases across some defining characteristics**. By doing this during the initial stages of evaluation design and in coordination with the implementing organization, the evaluator can share feedback on their suggested approach, if any cases need to be dropped, added, swapped, etc. to balance staying within scope, leveraging existing resources, and meeting evidence needs.

2. AVOID COMMON FACTOR DETERMINATION PITFALLS!

Selecting factors is similarly as important as selecting cases, and there are a few key ways that factor selection can go wrong. More details on the how-to steps for factor selection can be found in the Factor Selection section.

If you choose **factors that are too broad or mostly related to environmental factors**, you will struggle to limit the scope of the evaluation (given the broad nature of the factors) and you will have a harder time developing use-focused comparisons across the cases. For example, if the cases are intervention areas, and all the factors are broadly about the contextual environment, you will not be able to make any conclusions or recommendations on what intervention areas had successful results and why. To strike this balance in factor selection, you need to make sure you have a balance between environmental factors and programmatic or operational factors that you want to directly compare (i.e., implementation strategies, outcomes associated with intervention, etc.).

Relatedly, if you **leave out factors that are related to the results of the cases**, you will not have enough information to share with the implementing organization about how they may need to pivot their approach, what approaches are succeeding, or what conditions have enabled unexpected results.

Reviewing demographic or environmental factors should serve as a double-check that the cases you included are truly comparable. For example, with the red apples, you cannot make one of the factors "fruit type" if you are looking at different countries and there are no fruits they produce in common. There is not enough in common among those factors – how can you compare an orange on the West Coast of the United States to a mango in Thailand? What can you meaningful conclude about the differences in those cases beyond that they are different fruits, in different countries? To avoid comparisons like this, **make sure that you double down on your case selection and factor identification criteria**, ensuring that they will guide you to choose comparable cases with factors that explicitly and only answer the evaluation questions, which will lead to use-focused findings and conclusions for the implementing organization.

Lastly, **be careful not to choose too many factors.** QCA is distinct from other evaluation types, where adding an additional factor is not as simple as including one more question in an interview protocol. Each factor is associated with at least one question in the protocol, but often more than one. Additionally, as you add factors, you may need to also add in cases, to ensure comparability of the factor across the cases. As mentioned in the Scoping section above, each case you add will multiply the scale of the evaluation. This multiplier effect also demonstrates why it is so important to have these conversations with the implementing organization at the beginning of the design process, so they can understand the impact of what they may perceive as small changes or tweaks to the scope. Open and clear communication about the trade-offs and cascades of adding new factors will help ensure that you and the implementing organization are on the same page and can make adjustments or pivots together, acknowledging the potential increase in time, effort, budget, etc.

3. MAKE SURE YOU ACTUALLY HAVE ACCESS TO THE DATA YOU NEED!

As part of identifying and selecting the cases and the factors for your QCA evaluation, it is **essential to begin mapping out what data sources you will need to enable comparison across the factors, for all the cases.** These steps are shared in more detail in the Data Sources section, and a template for identifying data sources is included in Annex A. Once you have started to identify what the data sources are for each factor then you will need to assess whether collecting this data is feasible and if the data sources are accessible to you. If the answer is no, but the factors are essential to include in the evaluation, you will need to state these limitations and identify the mitigating actions you will take to access these data sources.

If those factors are more "nice to have" and not "need to have" for a successful evaluation, then you may consider dropping those factors, because not having good enough data sources will inhibit the analysis phase while still using up time and resources during data collection. If some factors do not have available data sources, it does not mean that you need to pick another evaluation method and start all over. Rather, **data availability is a consideration that may result in choosing different factors** and means raising concerns with your implementing organization in case they have access to some data sources that you do not. Conversations during design about data availability, alternative approaches, and limitations also help set expectations about the extent of findings or limited comparability issues ahead of the evaluation deliverables.

MODULE 6:

Examples of QCA

EXAMPLE A – USAID/ETHIOPIA STRATEGIC ADVISORY GROUP FOR EMERGENCIES (SAGE)/CRISIS MODIFIERS EVALUATION

Evaluation Report

Background: Headlight conducted a Developmental Evaluation (DE) with the Disaster Risk Management (DRM) and Effective Emergency Response (EER) Projects at USAID/Ethiopia. As part of the DE, the EER Project team determined that they wanted to learn more about the Mission's use of Crisis Modifiers (CMs), flexible funding for emergencies, specifically to explore if implementation has resulted in any substantiated outcomes and identify how the Mission can adapt this approach to respond to emergencies more effectively through development activities. Headlight began designing an evaluation to meet these needs in February 2022, starting by drafting four evaluation questions based on scoping conversations with the client. (page 5)

Selection of QCA: Once the evaluation questions were finalized, the Headlight team identified QCA as a best-fit method for this evaluation. The evaluation questions this evaluation sought to answer include:

1. To what extent is SAGE improving the coordination of EER?
2. What changes have resulted from the implementation of Crisis Modifiers?
3. How does SAGE engage with activities before, during application, and (if approved) after the Crisis Modifier has been used?
4. What are the types of Crisis Modifiers the Mission has used?

The objectives and questions agreed on by the client meant that the CMs would be the cases, and the cases had enough core components in common (i.e., all cases were CMs implemented in Ethiopia, funded by USAID, and occurred in a similar time window) to enable a QCA evaluation. Given the client's interest in the approach, process, implementation, and impacts of CMs, the Headlight team determined that cross-case comparisons, as is done through QCA, would yield actionable insights. Additionally, given the interest in outcomes from the CMs, the team also decided to embed OH within QCA for this evaluation to understand more about the CMs' results. (pages 6-7)

Implementation: First, the team worked with the client to prioritize the final set of cases. Using criteria from the client to include a diversity of geographies, implementing partners, and sectors, the team chose 13 cases (pages 8-10). Then, following the process used in Annex A (brainstorming, sorting, prioritizing, and identifying data sources), the team developed a list of 16 factors (page 6). The team then began sampling, interview instrument development, and data collection, where they conducted 176 key informant interviews (KIIs) and reviewed 57 secondary documents (pages 7, 11-12). During the coding and analysis phase, the team coded every KII and the secondary documents determined to have information on the factors. The team employed four analysis methods to analyze the factors: Process Analysis, Comparative Analysis, Narrative Analysis, and Thematic Inductive

Analysis. Lastly, the team produced a findings, conclusions, and recommendations (FCR) matrix and drafted several deliverables including a final report, a set of case studies (one per Crisis Modifier), an executive summary, and a presentation of the final FCR.

So What?: Using QCA enabled the evaluators to provide the client with actionable evidence on the conditions per case and recognize specific trends or commonalities among the cases. Employing qualitative analysis methods to analyze the factors allowed the evaluation team to provide nuanced information to the client regarding what does and does not work when implementing CMs, how enabling and inhibiting conditions such as conflict, insecurity, political will, etc. influenced the processes and outcomes of CMs and resulted in the identification of six substantiated outcomes and two emergent outcomes from CMs.

Now What?: With the information gathered by this evaluation and the resulting recommendations, the USAID/Ethiopia Mission and CM implementing partners are now working on identified actions to strengthen their emergency response skills, including using the SAGE platform for improved coordination and increased monitoring and measuring of outcomes from CMs. To build on this evaluation, the Headlight team further embedded a point of contact with an ongoing Crisis Modifier to identify learnings while the Activity is being implemented.

EXAMPLE B – HOUSEHOLD RESILIENCE DURING CONFLICT: QUALITATIVE COMPARATIVE ANALYSIS FOR THE CASE OF SYRIA

Evaluation Report

Background: In 2017, Mercy Corps conducted the Wages of War study to learn how Syrians cope with conflict and how they adapted their lives and livelihoods during the Syrian civil war. The Wages of War study relied on survey data and qualitative interviews with household members and community representatives in Syria. The study highlighted the range of complex factors that enabled and inhibited households from adapting livelihoods and improving welfare. Findings from this study were important to advance knowledge of how civilians cope during war, identify factors that can support resilience, and tailor humanitarian activities accordingly. The study looked at for Example B, "Household Resilience During Conflict: Qualitative Comparative Analysis for the Case of Syria" was conducted by the Resilience Evaluation, Analysis and Learning Associate Award and sought to leverage the data collected for the Wages of War study to investigate the methodological options for studying resilience in complex humanitarian settings.

Selection of QCA: The evaluators used fuzzy set QCA (fsQCA)[12] to study resilience due to the method's suitability for a small/medium purposively selected sample, its flexibility to use both quantitative and qualitative data, ability to assess complex causality, and the reliance on contextual knowledge which supports a participatory approach to research (pages 4-5). Specifically, the evaluation questions include, "*what factors, if reinforced, have the greatest potential to strengthen Syrian households' ability to adapt their lives and livelihoods, and cope with the crisis?"*, and additional questions are specific to the fsQCA method and its utility to study resilience in humanitarian contexts (page 5). Additional details on why QCA was chosen for this evaluation are further documented in the report (pages 6-18).

Implementation: The evaluators identified three resilience-related outcomes of interest, seven causal conditions theorized to connect to the outcomes, and two main macro-conditions (a favorable economic situation and good governance), which collectively composed the evaluation's factors. Cases were selected, *"from members of households and community key informants in three regions within Syria—northeast, north, and south-central Syria. "In total, representatives from 1,168 households and 350 key informant interviews were surveyed in 124 communities,"* (page 15).

The evaluators then ran several analyses per outcome using the existing quantitative data from household surveys collected during the Wages of War study. The analyses used different threshold settings to ensure that no threshold was arbitrarily selected and to identify causal relations stemming from singular or multiple factors (pages 15-18). From this analysis, the evaluators produced a set of findings for the first two outcomes, sharing the causal conditions and pathways to reach these outcomes, at different threshold levels (pages 18-22).

So What?: The evaluators concluded that applying the fuzzy set QCA method to the Wages of War dataset demonstrated the method's success at identifying complex concepts related to the resilience and welfare of Syrians affected by the civil war. The findings in the report highlight connections between the resilience outcomes and causal conditions, such as households that have not been exposed to intense conflict were found to have avoided extreme household hunger. The findings

12 "There are two main types of Qualitative Comparative Analysis—crisp set QCA (csQCA) and fuzzy-set QCA (fsQCA). CsQCA relies on the dichotomous definition of outcomes and conditions, taking the value of either 0 or 1. FsQCA, in contrast, allows for more gradation, where a case (e.g. member of a sample) will belong to a range of sets (variables) to varying degrees (assigned a numerical value ranging from 0 to 1)." (page 6, Qualitative Comparative Analysis for the Case of Syria)

also showed how the conditions exist together or separately in resilience households; for example, *"households who were located in communities marked by good governance AND had in-house female or youth income earners were found to have successfully avoided extreme household hunger,"* (page 19). The evaluators concluded that *"showing the relationships between diverse conditions allow for immediate programmatic recommendations,"* (page 22).

Now What?: The evaluators identified QCA as a good fit for future resilience evaluations and research as well as highlighted the next steps to further investigate the causal mechanisms that connect the causal conditions to the outcomes, such as conducting qualitative case studies (page 23).

EXAMINING IN THE DIFFERENT EXAMPLES

Despite both evaluations leveraging the QCA method as an overarching evaluation structure, they contextualized and approached several aspects of this method differently. The table below presents a brief comparison of method application differences between each evaluation to help the reader understand the spectrum through which the method can be applied based on practical examples.

Comparison of QCA Contextualization & Application

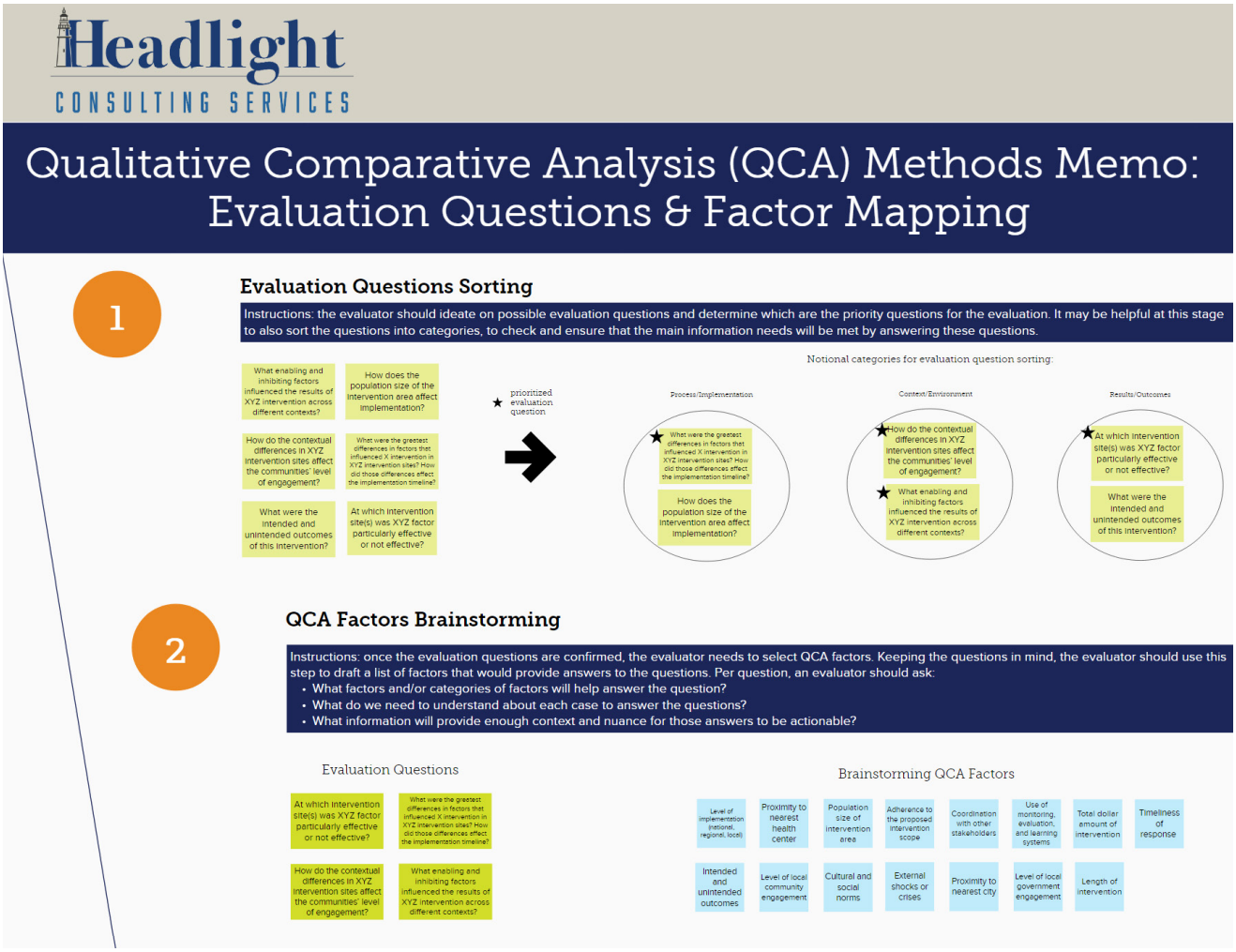
	Example A	Example B
Using QCA to determine causality	The overarching evaluation approach did look at causality, but through the integration of Outcome Harvesting as a secondary evaluation method best suited to this type of evaluation question. Causal mechanisms were assessed as part of verifying how outcomes occurred, if at all. (Pages 28 -32 in the evaluation)	The fuzzy set QCA approach was used to identify which conditions need to be present or absent for an outcome to occur. The report acknowledges there are multiple causal pathways for a given outcome, and, "social phenomena, including resilience, are theorized to be influenced by a variety and confluence of diverse factors, and as such, fsQCA may be an ideal method to unpack this complexity," (Page 6 of the evaluation).
Emphasis on qualitative versus quantitative data	There is a strong emphasis on qualitative information to understand how and why Crisis Modifiers are being used among USAID/Ethiopia activities and their effects. Using qualitative data from KIIs as well as secondary document review enabled the evaluators to identify the process, enabling/ inhibiting factors, and outcomes from USAID/Ethiopia activities using Crisis Modifiers. (evaluation question findings are presented throughout the main body of the report, beginning on page 22)	The focus was on quantitative data from household surveys on resilience, and the evaluators chose not to leverage qualitative data that they had collected from interviews. The data enabled the evaluators to share findings on the, "most salient combinations of causally relevant conditions (or pathways) that are linked to resilience," (page 18). The evaluators acknowledged that there are remaining questions on causal mechanism, stating that, "one potential way forward would be for qualitative researchers to identify the households that exhibit certain pathways, and then conduct interviews to unravel what the actual mechanisms are that link specific combinations of conditions to outcomes." (page 23)

These two different examples of QCA demonstrate that there are different ways to contextualize an evaluation method while upholding the core structure it provides to help answer relevant evaluation questions. Minimum standards help ensure the intended structure of a method is well-grounded and serves its intended purpose, and as long as evaluators follow the minimum standards and have selected the right method based on the evaluation questions, then there is a larger degree of flexibility in method implementation, from the number of cases selected to the emphasis on quantitative or qualitative date, etc.

Annex A:

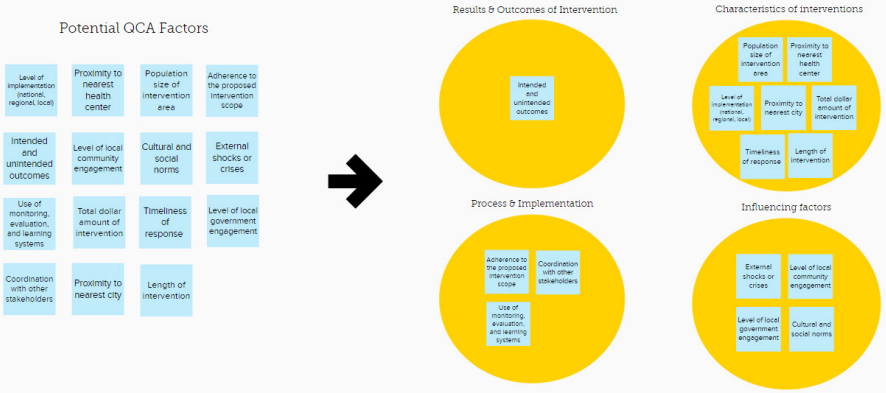
QCA Process Template

We have put together a template for evaluators to use when designing their QCA evaluation as well as an anonymized example (also included below) so that evaluators can visualize what it looks like when used. The template focuses on the preliminary phases of QCA design, including brainstorming evaluation questions and factors, sorting and prioritizing the factors according to evaluation question alignment, finalizing the factors, and beginning to identify data sources.



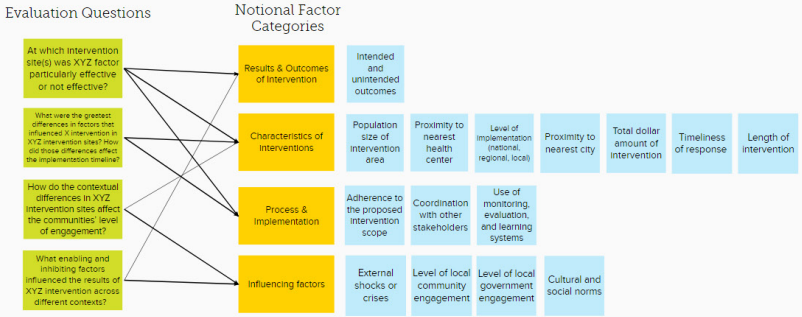
3 QCA Factors Sorting

Instructions: with the list of possible factors drafted, the evaluator should sort them into categories, similar to the first step with evaluation questions. These categories can be derived from the evaluation questions and anticipated areas for analysis.



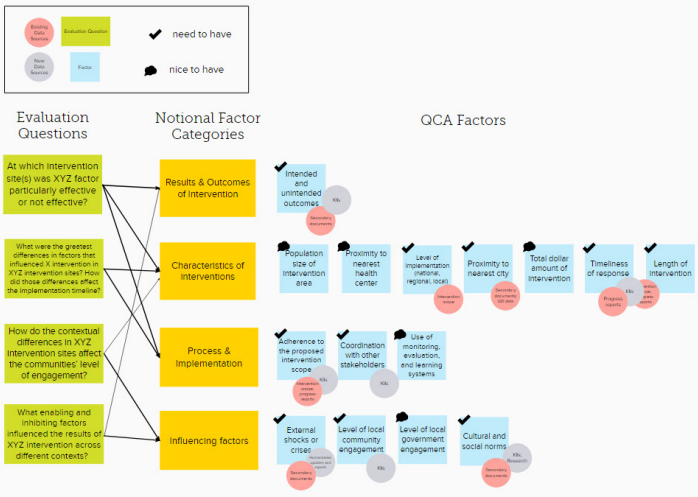
4 QCA Factors Mapping

Instructions: once the factors are sorted into categories, map each evaluation question to the relevant factor categories. At this stage, the evaluator should confirm that each evaluation question is mapped to at least one factor category, thus ensuring that at least one factor is designed to have answers for that evaluation question.



5 QCA Factors Finalization & Identification of Data Sources

Instructions: here the evaluator should go through the potential factors and determine which are necessary to answer the evaluation question and which may be merely nice to know, which are specific enough for comparison, and confirm that there are enough factors to answer the evaluation questions. Also in this step, the evaluator needs to identify the data sources for information on each factor (e.g., secondary documents, key informant interviews, etc.) and it is useful to indicate what data the evaluator already has versus what they need to collect.



Annex B:

Evaluator Terms of Reference

Organization Background

Provide an overview of the implementing organization, complementing the project background.

Relevant Project Background

This section should include a description of the problem the project seeks to address and the project's narrative or visual theory of change.

Scope of Work:

Provide an overview of the intended scope of the evaluative effort. What are the core expected tasks for the consultant or evaluation team? Will they be responsible for design, implementation, and adaptation support? Who will they be primarily working with at the implementing organization? Who do they need to coordinate with? How much support are they expected to provide outside the data collection and analysis? A very brief example of a scope overview is provided below.

The consultant will design and implement a Qualitative Comparative Analysis evaluation to capture the results of the activity/program/strategy in question. This work will include narrowing down the number of preliminarily identified cases to a use-focused number and ensuring comparability; selecting the factors for Comparison Analysis, informed by the overarching evaluation questions and consultations with the implementing organization; collecting data that will enable comparison of the factors among cases; and conducting analysis on all the factors to identify what factors most lead to success, how certain factors interact with each other, what inhibiting environment conditions should be mitigated against in the future, etc. From this analysis, the evaluator should prepare a final report that includes recommendations for how the evidence can be used to adapt implementation and further learning.

Expected Period of Performance

Provide the length of time the evaluative effort and support is expected to last, as well as any other engagement implications (such as regular check-ins, dissemination workshop findings, after action reviews, etc.). As a reminder, Qualitative Comparative Analysis is best used as a summative method and should be given at least six months to be conducted from design to dissemination after contracting, but that timeline varies depending on the total number of cases selected.

Resources

Provide an overview of the funds and/or level of effort anticipated for the evaluative effort.

Anticipated Tasks & Responsibilities:

Make sure to include a more detailed description of the anticipated tasks and responsibilities for the consultant so that they can appropriately match skills, knowledge, and availability with the desired service delivery. This section should include preliminary identification of cases so the evaluator has a sense of the scope/scale of the evaluative needs, the level of collaboration desired throughout, and any special support needs. This is also an excellent section to include more details about the amount of support needed after the evaluation is complete. The implementing organization should highlight what types of workshops or sessions they would like to disseminate the evidence and how much facilitation support is desired for adaptations.

Deliverables:

Include any early expectations around the type of deliverables that are needed and will be most useful from the evaluation. This should include any evaluation design deliverables (i.e., inception report, sampling strategy, interview protocol, etc.), final deliverables, and support for all necessary audiences. The implementing organization should specify if they want just one final summative comparison report or if they would like a comprehensive final report in addition to individual case studies and any specific length and formatting guidelines.

Qualifications:

It is also valuable to include qualifications to ensure appropriate candidates submit proposals. We have provided some sample, QCA-relevant qualifications below.

- Master's Degree and 8 years' relevant experience* OR Ph.D. and 6 years' experience;
- 6 years' experience in international development;
- Experience with [INSERT DONOR];
- Knowledge of qualitative evaluation methods (experience implementing QCA preferable);
- Experience implementing various analysis methods (e.g., Thematic Inductive, Thematic Deductive, Process, Narrative, Comparative, etc.);
- Ability to contextualize and adapt methods;
- Experience conducting semi-structured interviews;
- Ability to gather evidence and present analysis in a competent and engaging manner;
- Excellent written and oral communication skills;
- Ability to work collaboratively within and across teams;
- Experience in examining large amounts of data and extracting relevant information for a summary; and,
- Experience in stakeholder engagement and facilitation.

**While number of years experience is a flawed metric, the success of QCA depends on how many different qualitative methods an evaluator has been exposed to, if they can run semi-structured interviews, and understanding of comparative experiences. The more time a person has spent in the profession means that they are more likely to be exposed to different components that they can use for implementation reference.*

Annex C:
Bibliography

David L. Cooperrider Center for Appreciative Inquiry, Champlain College | Stiller School of Business. (2023, April 7). An Introductory Guide to Appreciative Inquiry. https://appreciativeinquiry.champlain.edu/2023/04/07/introduction-to-appreciative-inquiry/

Dedoose. (n.d.). Dedoose User Guide. Retrieved December 8, 2023 from https://www.dedoose.com/userguide

Guest, G. (2017, March 29). A pathway for sampling success. R&E Search for Evidence. https://researchforevidence.fhi360.org/pathway-sampling-success

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. Field Methods, 18(1), 59–82. https://doi.org/10.1177/1525822X05279903

Howe, K., Krystalli, R., Krishnan, V., Kurtz, J., & Macaranas, R. ((2018, February 20). The Wages of War: Learning from how Syrians have adapted their livelihoods through seven years of conflict. MercyCorps. https://www.mercycorps.org/research-resources/wages-war

Howe, K. & Patros, T. (2020). Household Resilience During Conflict: Qualitative Comparative Analysis for the Case of Syria. The Resilience Evaluation, Analysis and Learning (REAL) Associate Award. https://pdf.usaid.gov/pdf_docs/PA00X3DZ.pdf

INTRAC. (2017). Process Tracing. https://www.intrac.org/resources/process-tracing/

Kuhn, C. (2020, August 17). Choosing the Right Qualitative Analysis Software. Headlight Consulting Services. https://www.headlightconsultingservices.com/blog/choosing-the-right-qualitative-analysis-software/

Kuhn, C. (2020, September 1). Top 5 Dos and Don'ts of Qualitative Analysis. Headlight Consulting Services. https://www.headlightconsultingservices.com/blog/top-5-dos-and-donts-of-qualitative-analysis/

Kuhn, C. (2020, September 16). Findings, Conclusions, and Recommendations Matrices. Headlight Consulting Services. https://www.headlightconsultingservices.com/blog/findings-conclusions-and-recommendations-matrices/

Kuhn, C., Harrell, A., Woldeeyesus, E., & Herrington, R. (2022). SAGE/Crisis Modifiers Evaluation Final Report. Headlight Consulting Services. https://www.headlightconsultingservices.com/resources/sage-crisis-modifiers-evaluation-final-report-june-2022/

Kuhn, C. & Herrington, R. (2022, May 5). Outcome Harvesting Methods Memo. Headlight Consulting Services. https://www.headlightconsultingservices.com/resources/outcome-harvesting-methods-memo/

Marx, A., Rihoux, B., & Ragin, C. (2014). The origins, development, and application of Qualitative Comparative Analysis: The first 25 years. European Political Science Review, 6(1), 115-142. doi:10.1017/S1755773912000318

Namey, E. (2017, April 25). Riddle me this: How many interviews (or focus groups) are enough? R&E Search for Evidence. https://researchforevidence.fhi360.org/riddle-me-this-how-many-interviews-or-focus-groups-are-enough

Positive Deviance. (2021, November 3). BetterEvaluation. https://www.betterevaluation.org/en/plan/approach/positive_deviance

Ragin, C. (1984). What is Qualitative Comparative Analysis (QCA)? Department of Sociology and Department of Political Science, University of Arizona. https://eprints.ncrm.ac.uk/id/eprint/250/1/What_is_QCA.pdf

Secskas, M. & Kuhn, C. (2022, March 10). Top 5 More Dos and Don'ts of Qualitative Analysis. Headlight Consulting Services. https://www.headlightconsultingservices.com/blog/top-5-more-dos-and-donts-of-qualitative-analysis/

Sheth, K. (2018, August 16). Top Apple Producing Countries In The World. World Atlas. https://www.worldatlas.com/articles/top-apple-producing-countries-in-the-world.html#:~:text=Apple%20is%20one%20of%20the%20most%20produced%20fruits%20in%20the,44%20million%20tonnes%20produced%20annually.

Simister, N., & Scholz, V. (2017). QUALITATIVE COMPARATIVE ANALYSIS (QCA). INTRAC. https://www.intrac.org/wpcms/wp-content/uploads/2017/01/Qualitative-comparative-analysis.pdf

UNICEF. (n.d.). Randomized Control Trials. Retrieved December 8, 2023 from https://www.unicef-irc.org/KM/IE/impact_7.php#:~:text=A%20randomized%20controlled%20trial%20

USAID. (n.d.). Rapid Feedback MERL. Retrieved December 8, 2023 from https://www.usaid.gov/PPL/MERLIN/RapidFeedbackMERL

USAID. (2017). Shock Responsive Programming and Adaptive Mechanisms: Approaches and Tools. USAID Learning Lab. https://usaidlearninglab.org/sites/default/files/resource/files/shock_responsive_programming_guidance_compliant.pdf

Wilson-Grau, R., & Britt, H. (2012). Outcome Harvesting (p. 31). Ford Foundation. https://outcomeharvesting.net/outcome-harvesting-brief/