BetterEvaluation

Manage an evaluation or evaluation system

Manage an evaluation (or a series of evaluations), including deciding who will conduct the evaluation and who will make decisions about it.

1. Understand and engage with stakeholders

Who needs to be involved in the evaluation? How can they be identified and engaged?

Understand stakeholders

Community scoping: developing a more in-depth understanding of a community of interest by providing information about its social diversity, history, existing networks, and overall socio-economic characteristics.

Stakeholder mapping and analysis: identifying different stakeholders' level of interest and influence.

Understanding and taking into account the priorities and concerns of different stakeholders informs evaluation planning, communication strategies during and after the evaluation and supports the utilization of evaluation findings.

Engage stakeholders

Community fairs: organising a community event with the aim of providing information about the project and raising the awareness of relevant issues.

Fishbowl technique: managing group discussion about relevant issues.

Formal meeting processes: guidance on processes for running formal meetings.

Informal meeting processes: a conversation between an evaluator and a key stakeholder that is not conducted in a formal way but is still seeking the same outcomes.

2. Establish decision making processes

Who will have the authority to make what type of decisions about the evaluation? Who will provide advice or make recommendations about the evaluation? What processes will be used for making decisions?

Types of structures

Advisory group: forming a group to provide advice on evaluations without making any actual decisions.

Citizen Juries: using representatives from the wider community to make decisions about possible approaches or options.

Steering group: establishing a group to make decisions about an evaluation.

Ways of operating

Consensus decision making: decision making processes that aims to find decisions which everyone can accept.

Formal meeting processes: guidance on processes for running formal meetings.

Hierarchical decision making: making decisions on the basis of formal positions of authority.

Informal meeting processes: a conversation between an evaluator and a key stakeholder that is not conducted in a formal way but is still seeking the same outcomes.

Majority decision making: basing decisions which have the support of the majority of the decision makers.

Round robin: a structured process for generating ideas in a group.

Six Hats Thinking: promoting holistic and lateral thinking in decision-making and evaluation by using different roles.

Approaches

Participatory evaluation: involving key stakeholders in the evaluation process.

3. Decide who will conduct the evaluation

Who will actually conduct the evaluation?

Community: conducting an evaluation by using the broader community or groups of intended beneficiaries.

Expert Review: conducting an evaluation by using someone with specific content knowledge or expert judgment and professional expertise.

External Consultant: contracting an external consultant to conduct the evaluation.

Hybrid - Internal and External Evaluation: a combination of internal staff and an external (usually expert) opinion to jointly conduct an evaluation.

Internal Staff: conducting an evaluation using staff from the implementing agency.

Learning Alliances: bringing together different groups to conduct the evaluation.

Peer Review: conducting an evaluation using individuals/organizations who are working on similar projects.

Approaches

Horizontal Evaluation: conducting an evaluation through a structured approach to peer learning.

4. Determine and secure resources

What resources (time, money, and expertise) will be needed for the evaluation and how can they be obtained? Consider both internal (e.g. staff time) and external (e.g. previous participants' time).

Determine resources needed:

Evaluation budget matrix: creating a budget using an evaluation budget matrix.

Evaluation costing: calculating the costs of an evaluation including time, money and expertise.

Resources stocktake: stocktaking resources for an evaluation which can include people's time and expertise, equipment and funding.

Secure resources needed:

Designated staff time for evaluation: at the proposal or planning stage, ensuring that staff have time to conduct, participate in or reflect on the results of evaluations as part of organizational learning.

Grant funding for evaluation: writing evaluation in to grants and/or writing a separate grant to request funding for an evaluation.

Institutionalized budget allocation: having a policy where funding allocation for evaluation is built into project and program budgets (either a fixed amount or a percentage of the total budget).

Reduced evaluation costs: reducing the costs is an option to consider if evaluation costs outweigh the predicted benefits.

5. Define ethical and quality evaluation standards

What will be considered a high quality and ethical evaluation? How should ethical issues be addressed?

Cultural Competency: ensuring the influence of culture on human behaviour is taken into consideration during the evaluation.

Ethical guidelines: institutional or organizational rules or norms that guide evaluation practice, especially regarding vulnerable populations.

Evaluation standards: core national or internationally agreed best practice for conducting evaluation.

Institutional review board: a committee set up by an organization or institution to monitor the ethical and technical research and evaluation conducted by its members.

6. Document management processes and agreements

How will you document the evaluation's management processes and agreements made?

Document setting out what is needed in an evaluation:

Terms of Reference: sets out the purpose(s) of the evaluation and its key evaluation questions, as well as a timeline (and possibly the available budget and existing data resources); for external evaluations this can include contractual arrangements. In some cases this includes an evaluation plan setting out the methodology.

Request for Proposals (RFP): RFPs (also known as Request for Application or Request for Quotation) are formal request for evaluators to prepare a response to a planned evaluation and are generally used to select the final evaluator for the evaluation.

Document setting out how different organizations will work together:

Contractual Agreement: creating formal contracts to engage external evaluators.

Memorandum of Understanding: high level agreement between two or more organizations committing to work together.

7. Develop an evaluation plan, framework or policy

What is the overall plan for the evaluation? Is there a larger evaluation framework across several related evaluations?

Evaluation framework: provides an overall framework across different evaluations, such as individual projects within a program.

Evaluation plan: set out the proposed details of an evaluation including what, how and when an evaluation will take place.

Evaluation policy: processes, structures and principles which will guide individual evaluation and the development of evaluation capacity within an organization.

8. Review evaluation (do meta-evaluation)

How will evaluation reports be reviewed before they are finalized? Will there be a review of the evaluation process to improve this?

Beneficiary exchange: seeking feedback from the beneficiaries by discussing the findings with them.

Expert review: reviewing the evaluation by using subject experts, either individually or as a panel.

External review: garnering feedback from external experts or anonymous reviewers.

Group critical reflection: facilitating a group stakeholder feedback session.

Individual critical reflection: asking particular individual stakeholders for their independent feedback.

Peer review: reviewing the evaluation by using peers from within the organisation or outside of the organisation.

9. Develop evaluation capacity

How can the ability of individuals, groups and organizations to conduct and use evaluations be strengthened?

Conferences: attendance at professional conferences to understand how other evaluators frame and discuss their findings is a key component of building evaluation capacity.

Mentoring: supporting a colleague by sharing professional and personal experiences in order to support their development and growth.

Peer review: reviewing the evaluation by using peers from within the organisation or outside of the organisation.

Training and formal education: developing people's knowledge and skills in conducting and/or managing an evaluation.

Find options (methods), resources and more information on these tasks and approaches online at http://betterevaluation.org/plan/manage_evaluation

BetterEvaluation

Define what is to be evaluated

Develop a description (or access an existing version) of what is to be evaluated and how it is understood to work.

1. Develop initial description

What exactly is being evaluated?

Peak Experience Description: describing a time when the project/program/policy worked particularly well. This option is part of the Appreciative Inquiry approach to evaluation.

Thumbnail Description: briefly describing the project/program/policy.

Approaches

Appreciative Inquiry: uses a group process which allows an organisation to understand the best elements of itself in order to create a better future.

2. Develop program theory/logic model

How is the intervention (project, program, policy, etc.) understood to work (program theory, theory of change, logic model)?

Ways of developing logic models:

Articulating Mental Models: talking individually or in groups with key informants (including program planners, service implementors and clients) about how they understand an intervention works.

Backcasting: working backward from a desirable future, to the present in order to determine the feasibility of the idea or project.

Five Whys: asking questions in order to examine the cause-and-effect relationships that create underlying problems.

Group Model Building: building a logic model in a group, often using sticky notes.

Previous Research and Evaluation: extracting information about intended and actual outcomes and impacts and important aspects of context and implementation from previous reports.

SWOT Analysis: reflecting on and assessing the Strengths, Weaknesses, Opportunities and Threats of a particular strategy in order to discover how it can best be implemented.

Tiny Tools Results Chain: mapping both positive and negative possible impacts from an intervention.

Ways of representing logic models:

Logframe: designing, executing and assessing projects by considering the relationships between available resources, planned activities, and desired changes or results.

Outcomes Hierarchy: (also known as a theory of change or an outcomes chain) demonstrating a series of outcomes leading up to the final impacts of a project.

Realist Matrix: focusing on one of the steps in an outcomes chain and then identifying the mechanism involved in producing the outcome and the contexts within which this mechanism operates.

Results Chain: (also known as a pipeline model) showing a program as a series of boxes - inputs-> activities -> outputs -> outcomes -> impacts.

See the next page for two examples of representing a logic model: using an Outcomes Hierarchy and a Results Chain.

Approaches

Collaborative Outcomes Reporting: uses a collaborative approach to developing a logic model.

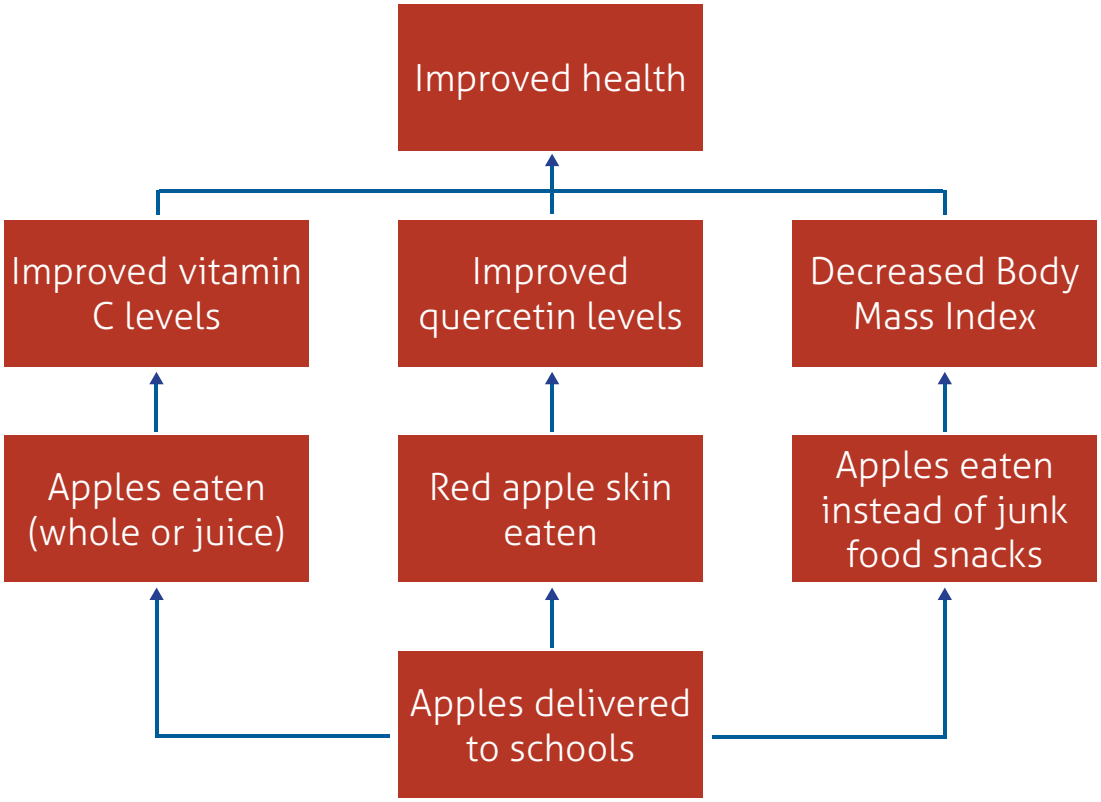
Outcome Mapping: focuses on identifying 'boundary partners' whose work is not under the control of a program but who are critically important to its success in achieving its objectives.

Participatory Impact Pathways Approach: focuses particularly on identifying the networks of actors involved and how these networks need to change in order for the objectives to be achieved.

Different ways of presenting logic models:

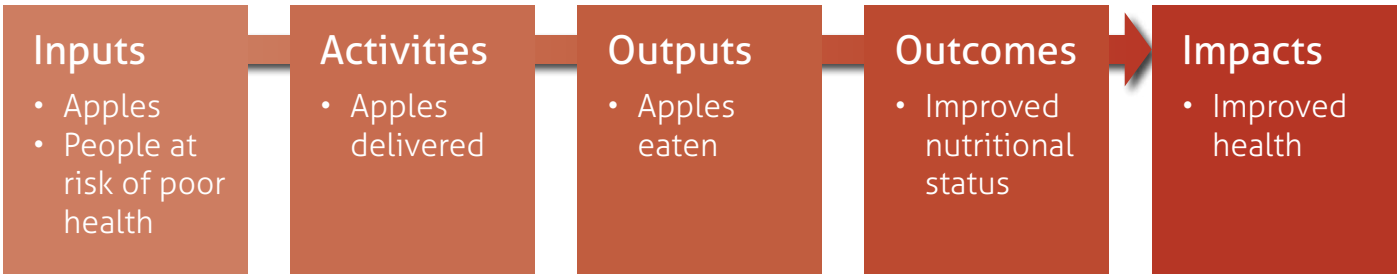
Outcomes Hierarchy (also known as theory of change or an outcomes chain)

demonstrating a series of outcomes leading up to the final impacts of a project, sometimes including different possible causal pathways.



Results Chain (also known as a pipeline model)

showing a program as a series of boxes - inputs-> activities -> outputs -> outcomes -> impacts.



3. Identify potential unintended results

What are possible unintended results (both positive and negative) that will be important to address in the evaluation?

Key Informant Interviews: asking experienced people to identify possible negative impacts, based on their experience with similar programs. Program critics can be especially useful.

Negative Program Theory: identifying ways in which program activities might produce negative impacts rather than their intended impacts.

Risk Assessment: identifying the potential negative impacts, their likelihood of occurring and how they might be avoided.

Six Hats Thinking: promoting holistic and lateral thinking in decision-making and evaluation.

Find options (methods), resources and more information on these tasks and approaches online at http://betterevaluation.org/plan/define

BetterEvaluation

Frame the boundaries for an evaluation

Set the parameters of the evaluation – its purposes, key evaluation questions and the criteria and standards to be used.

1. Identify primary intended users

Who are the primary intended users of this evaluation?

This task does not have specific options but does have resources to help guide you.

2. Decide purposes

What are the primary purposes and intended uses of the evaluation?

Using findings:

Contribute to broader evidence base: inform future policy and practice by others outside the organization.

Inform decision making aimed at improvement (formative): changing or confirming policies and practices.

Inform decision making aimed at selection, continuation or termination (summative): identifying best value for money.

Lobby and advocate: justify expenditure and demonstrate achievements.

Using process:

Build trust and legitimacy across stakeholders: develop better understandings of each other and demonstrate that expectations are being met.

Ensure accountability: holding someone to account to someone for something.

Ensure diverse perspectives are included, especially those with little voice: make explicit the experiences and values of key stakeholders, especially intended beneficiaries.

3. Specify the key evaluation questions

What are the high level questions the evaluation will seek to answer? How can these be developed?

This task does not have specific options but does have resources to help guide you. In addition, be clear about the different types of questions you want the evaluation to answer:

Descriptive question - what has happened? what is the situation?

For example - Where has the program been delivered? What changes have occurred for participants?

Causal question – what caused or contributed to the results?

For example - What were the outcomes and impacts of the program? What other factors contributed to achieving these outcomes and impacts?

Synthesis question – is this good? In what ways could it be better? Is it the best option?

For example -Did service delivery comply with agreed standards? Was the program cost-effective? What were its strengths and weaknesses?

Action question – what action should be taken?

For example -Should the program continue? What changes should be made to the program? Should it be scaled up?

4. Determine what success looks like

What should be the criteria and standards for judging performance? Whose criteria and standards matter? What process should be used to develop agreement about these?

Formal statements of values:

DAC Criteria: setting out high level evaluation criteria for evaluations which must be operationalized for each evaluation (OECD's Development Assistance Committee).

Millenium Development Goals (MDGs): a set of time bound and quantified goals and targets developed to help track progress in eradicating poverty

Standards, evaluative criteria and benchmarks: developing explicit standards, evaluative criteria or benchmarks or using existing relevant standards, criteria or benchmarks to define values.

Stated goals and objectives (including legislative review and policy statements): stating the program's objectives and goals so they can be used to assess program success.

Articulate and document tacit values:

Hierarchical Card Sorting (HCS): a participatory card sorting option designed to provide insight into how people categorize and rank different phenomena.

Open space technology: facilitating a group of 5 - 500 people in which a central purpose, issue, or task is addressed without a formal initial agenda.

Photovoice: using cameras to allow participants (often intended beneficiaries) to take and share photos in order to describe how they relate to important issues for them.

Rich Pictures: exploring, acknowledging and defining a situation through diagrams in order to create a preliminary mental model.

Stories of change: showing what is valued through the use of specific narratives of events.

Values Clarification Interviews: interviewing key informants and intended beneficiaries to identify what they value.

Values clarification public opinion questionnaires: seeking feedback from feedback from large numbers of people about their priorities through the use of questionnaires.

Negotiate between different values:

Concept Mapping: negotiating values in order to frame the evaluation.

Delphi Study: generating a consensus without face to face contact by soliciting opinions from individuals in an iterative process of answering questions.

Dotmocracy: recording participants opinions by using sticky dots to either record agreement or disagreement with written statements.

Open Space Technology: facilitating a group of 5 - 500 people in which a central purpose, issue, or task is addressed without a formal initial agenda.

Public Consultations: conducting public meetings to provide an opportunity for the community to raise issues of concern and respond to options.

Approaches

Critical System Heuristics: an approach used to surface, elaborate, and critically consider boundary judgments, that is, the ways in which people/groups decide what is relevant to the system of interest (any situation of concern).

Find options (methods), resources and more information on these tasks and approaches online at http://betterevaluation.org/plan/engage_frame

BetterEvaluation

Describe activities, outcomes, impacts and context

Collect and retrieve data to answer descriptive questions about the activities of the project/program/policy, the various results it has had, and the context in which it has been implemented.

1. Sample

What sampling strategies will you use for collecting data?

Probability:

Multi-Stage: cluster sampling in which larger clusters are further subdivided into smaller, more targeted groupings for the purposes of surveying.

Sequential: selecting every n^{th} case from a list (e.g. every 10^{th} client).

Simple Random: drawing a sample from the population completely at random.

Stratified Random: splitting the population into strata (sections or segments) in order to ensure distinct categories are adequately represented before selecting a random sample from each.

Purposeful:

Confirming and Disconfirming: providing deeper insights into preliminary findings and highlighting the boundaries of the findings.

Criterion: involving the identification of particular criterion of importance, the articulation of these criterion, and the systematic review and study of cases that meet the criterion.

Critical Case: identifying cases that have the potential to impact other cases.

Homogenous: selecting similar cases to further investigate a particular phenomenon or subgroup of interest.

Intensity: selecting cases which exhibit a particular phenomenon intensely.

Maximum Variation: contains cases that are purposefully as different from each other as possible.

Outlier: analysing cases that are unusual or special in some way, such as outstanding successes or notable failures.

Snowball: asking a number of people where else to seek information creates a snowball effect as the sample gets bigger and bigger and new information rich examples are accumulated

Theory-based: selecting cases according to the extent to which they represent a particular theoretical construct.

Typical Case: developing a profile of what is agreed as average, or normal.

Convenience:

Convenience: based on the ease or "convenience" of gaining access to a sample. simply in which data is gathered from people who are readily available.

Volunteer: sampling by simply asking for volunteers

2. Use measures, indicators or metrics

What measures or indicators will be used? Are there existing ones that should be used or will you need to develop new measures and indicators?

This task has resources only in the areas of:

- Gender issues
- Governance
- Health
- Human rights
- Inequality
- Poverty
- Quality of life
- Wellbeing.

3. Collect and/or retrieve data

How will you collect and/or retrieve data about activities, results, context and other factors?

Individuals:

Convergent Interviewing: asking probing questions to interviewees and then using reflective prompts and active listening to ensure the conversation continues.

Deliberative Opinion Polls: providing information about the issue to respondents to ensure their opinions are better informed.

Email Questionnaires: distributing questionnaires online via email.

Face to Face Questionnaires: administering questionnaires in real time by a researcher reading the questions (either face to face or by telephone).

Global Assessment Scales: providing an overall rating of performance across multiple dimensions (also called a rubric).

Goal Attainment Scales: recording actual performance compared to expected performance using a 5 point scale from -2 (much less than expected) to +2 (much more than expected).

Internet Questionnaires: collecting data via a form (with closed or open questions) on the web.

Interviews: in-depth, structured, semi-structured, or unstructured.

Key Informant Interviews: interviewing people who have particularly informed perspectives.

Logs and Diaries: monitoring tools for recording data over a long period of time.

Mobile Phone Logging: Targeted gathering of structured information using devices such as smartphones, PDAs, or tablets.

Peer/Expert Reviews: Drawing upon peers or experts with relevant experience and expertise to assist in the evaluation of some aspect or all of a project.

Photo Voice: promoting participatory photography as an empowering option of digital storytelling for vulnerable populations.

Photolanguage: eliciting rich verbal data where participants choose an existing photograph as a metaphor and then discuss it.

Polling Booth: collect sensitive information from participants anonymously .

Postcards: collecting information quickly in order to provide short reports on evaluation findings (or an update on progress).

Projective Techniques (photo-elicitation): participants selecting one or two pictures from a set and using them to illustrate their comments about something.

Seasonal Calendars: analysing time-related cyclical changes in data.

Sketch Mapping: creating visual representations ('map') of a geographically based or defined issue.

Stories (Anecdote): providing a glimpse into how people experience their lives and the impact of specific projects/programs.

Survey: collecting data in response to structured questions.

Telephone Questionnaires: administering questionnaires by telephone.

Groups:

After Action Review: bringing together a team to discuss a task, event, activity or project, in an open and honest fashion.

Brainstorming: focusing on a problem and then allowing participants to come up with as many solutions as possible.

Card Visualization: brainstorming in a group using individual paper cards to express participants thoughts about particular ideas or issues.

Concept Mapping: showing how different ideas relate to each other - sometimes this is called a mind map or a cluster map.

Delphi Study: soliciting opinions from groups in an iterative process of answering questions in order to gain a consensus.

Dotmocracy: collecting and recognizing levels of agreement on written statements among a large number of people.

Fishbowl Technique: managing group discussion by using a small group of participants to discuss an issue while the rest of the participants observe without interrupting.

Focus Groups: discovering the issues that are of most concern for a community or group when little or no information is available.

Future Search Conference: identifying a shared vision of the future by conducting a conference with this as its focus.

Hierarchical Card Sorting: a participatory card sorting option designed to provide insight into how people categorize and rank different phenomena.

Keypad technology: gauging audience response to presentations and ideas in order to gain provide valuable feedback from large group settings.

Mural: collecting data from a group of people about a current situation, their experiences using a service, or their perspectives on the outcomes of a project.

ORID: enabling a focused conversation by allowing participants to consider all that is known (Objective) and their feelings (Reflective) before considering issues (Interpretive) and decisions (Decisional).

Q-methodology: investigating the different perspectives of participants on an issue by ranking and sorting a series of statements (also known as Q-sort).

SWOT Analysis: reflecting on and assessing the Strengths, Weaknesses, Opportunities and Threats of a particular strategy.

World Cafe: hosting group dialogue in which the power of simple conversation is emphasised in the consideration of relevant questions and themes.

Writeshop: a writing workshop involving a concentrated process of drafting, presenting, reviewing and revising documentations of practice

Observation:

Field Trips: organizing trips where participants visit physical sites.

Non-participant Observation: observing participants without actively participating.

Participant Observation: identifying the attitudes and operation of a community by living within its environs.

Photography/video: discerning changes that have taken place in the environment or activities of a community through the use of images taken over a period of time.

Transect: gathering spacial data on an area by observing people, surroundings and resources while walking around the area or community.

Physical measurements:

Biophysical: measuring physical changes over a period of time related to a specific indicator by using an accepted measurement procedure.

Geographical: capturing geographic information about persons or objects of interest such as the locations of high prevalence of a disease or the location of service delivery points.

Existing documents and data:

Big data: data sets that are so voluminous and from such different sources that traditional analysis methods are not feasible or appropriate.

Official Statistics: obtaining statistics published by government agencies or other public bodies such as international organizations. These include quantitative or qualitative information on all major areas of citizens' lives such as economic and social development, living conditions, health, education, the environment.

Previous Evaluations and Research: using the findings from evaluation and research studies that were previously conducted on the same or closely related areas.

Project Records: retrieving relevant information from a range of documents related to the management of a project such as the project description, strategic and work plans, budget and procurement documents, official correspondence, minutes of meetings, description and follow-up of project participants, progress reports.

Reputational Monitoring Dashboard: monitoring and quickly appraising reputational trends at a glance and from a variety of different sources.

4. Manage data

How will you organize and store data and ensure its quality?

Data Cleaning: detecting and removing (or correcting) errors and inconsistencies in a data set or database due to the corruption or inaccurate entry of the data.

Additional resources available on:

- Data management
- Data quality

5. Combine qualitative and quantitative data?

How will you combine qualitative and quantitative data?

When data are gathered:

Parallel Data Gathering: gathering qualitative and quantitative data at the same time.

Sequential Data Gathering (Sequencing): gathering one type of data first and then using this to inform the collection of the other type of data.

When data are combined:

Component Design: collecting data independently and then combining at the end for interpretation and conclusions.

Integrated Design: combining different options during the conduct of the evaluation to provide more insightful understandings.

Purpose of combining data:

Enriching: using qualitative work to identify issues or obtain information on variables not obtained by quantitative surveys.

Examining: generating hypotheses from qualitative work to be tested through the quantitative approach.

Explaining: using qualitative data to understand unanticipated results from quantitative data.

Triangulation (Confirming/reinforcing; Rejecting): verifying or rejecting results from quantitative data using qualitative data (or vice versa).

6. Analyze data

How will you investigate patterns in numeric or textual data?

Numeric analysis:

Correlation: a statistical technique to determine how strongly two or more variables are related.

Crosstabulations: getting an indication of the frequency of two variables (e.g. gender or income, and frequency of school attendance) occurring at the same time.

Data and text mining: computer-driven automated techniques that run through large amounts of text or data to find new patterns and information.

Exploratory Techniques: taking a 'first look' at a dataset by summarizing its main characteristics, often by using visual methods.

Frequency tables: arranging collected data values in ascending order of magnitude, along with their corresponding frequencies to ensure a clearer picture of a data set.

Measures of Central Tendency: a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution.

Measures of Dispersion: a summary measure that describes how values are distributed around the centre.

Multivariate descriptive: providing simple summaries of (large amounts of) information (or data) with two or more related variables.

Non-Parametric inferential: data that are flexible and do not follow a normal distribution.

Parametric inferential: carried out on data that follow certain parameters: the data will be normal (ie, the distribution parallels the bell curve); numbers can be added, subtracted, multiplied and divided; variances are equal when comparing two or more groups; and the sample should be large and randomly selected.

Summary statistics: providing a quick summary of data which is particularly useful for comparing one project to another, before and after.

Time series analysis: observing well-defined data items obtained through repeated measurements over time.

Textual analysis:

Content analysis: reducing large amounts of unstructured textual content into manageable data relevant to the (evaluation) research questions.

Thematic coding: recording or identifying passages of text or images that are linked by a common theme or idea allowing the indexation of text into categories.

7. Visualize data

How will you display data visually?

See relationships among data points:

Scatterplot: displaying the relationship between two quantitative variables plotted along two axes. A series of dots represent the position of observations from the data set.

Matrix Chart: summarising a multidimensional data set in a grid.

Network Diagram: a depiction of how people or other elements are related to one another.

Compare a set of values:

Bar Chart: illustrating the main features of the distribution of a data set in a clear way.

Block Histogram: presenting a frequency distribution of quantitative data in a graphical way.

Bubble Chart: providing a way to communicate complicated data sets quickly and easily.

Track rises and falls over time:

Line Graph: displaying information as a series of data points connected by straight line segments, on two axes.

Stacked Graph: visualising how a group of quantities changes over time. Items are "stacked" in this type of graph allowing the user to add up the underlying data points.

See the parts of a whole:

Pie Chart: a circular chart divided into sectors (like slices of a pie), illustrating proportion.

Treemap: makes use of qualitative information in the form of important distinctions or differences that people see in the world around them. They help overcome some of the problems that may be encountered when dealing with qualitative information.

Icon array: a matrix of icons (usually 100 or 1000 icons) typically used as a frequency-based representation of risk, simultaneously displaying both the number of expected events and the number of expected non-events.

Analyze a text:

Word Tree: a visual display of the words in qualitative dataset, where frequently used words are connected by branches to the other words that appear nearby in the data.

Phrase Net: depicts, in a network diagram, the relationships between different words in a source text using pattern matching (i.e., looks for pairs of words that fit a particular patterns). Matching different patterns provides different views of concepts contained in the text.

Word Cloud: assists an evaluator identify important words during the process of textual analysis.

See the world:

Demographic Mapping: using GIS (global information system) mapping technology to show data on population characteristics by region or geographic area.

Geotagging: adding geographic information about digital content, within "metadata" tags - including latitude and longitude coordinates, place names and/or other positional data.

GIS Mapping: creating very precise maps representing geographic coordinates that could include information relating to changes in geographical, social or agricultural indicators.

Interactive mapping: maps that allow users to interact – e.g. zooming in and out, panning around, identifying specific features, querying underlying data such as by topic or a specific indicator (e.g., socioeconomic status), generating reports

Social mapping: identifying households using pre-determined indicators that are based on socio-economic factors.

Find options (methods), resources and more information on these tasks and approaches online at http://betterevaluation.org/plan/describe

BetterEvaluation

Understand Causes of outcomes and impacts

Collect and analyze data to answer causal questions about what has produced outcomes and impacts that have been observed.

1. Check the results support causal attribution

How will you assess whether the results are consistent with the theory that the intervention produced them?

Gathering additional data:

Asking Key Informants to Attribute Causality: providing evidence that links participation plausibly with observed changes.

Modus Operandi: drawing on the previous experience of participants and stakeholders to determine what constellation or pattern of effects is typical for an initiative.

Process Tracing: focusing on the use of clues (causal-process observations, CPOs) to adjudicate between alternative possible explanations.

Analysis:

Check Dose-Response Patterns: examining the link between dose and response as part of determining whether the program caused the outcome.

Check Intermediate Outcomes: checking whether all cases that achieved the final impacts achieved the intermediate outcomes.

Check Results Match a Statistical Model: comparing results with a statistical model to determine if the program caused the outcome.

Check Results Match Expert Predictions: making predictions based on program theory or an emerging theory of wider contributors to outcomes and then following up these predictions over time.

Check Timing of Outcomes: checking predicated timing of events with the dates of actual changes and outcomes.

Comparative Case Studies: using a comparative case study to check variation in program implementation.

Qualitative Comparative Analysis: comparing the configurations of different cases to identify the components that produce specific outcomes.

Realist Analysis of Testable Hypotheses: Using a realist program theory (what works for whom in what circumstances through what causal mechanisms?) to identify specific contexts where results would and would not be expected and checking these.

Approaches: the following approaches combine some of the above options together with ruling out possible alternative explanations:

Contribution Analysis, Collaborative Outcomes Reporting, Multiple Lines and Levels of Evidence (MLLE), Rapid Outcomes Assessment. See below for definitions.

2. Compare results to the counterfactual

How will you compare the factual with the counterfactual - what would have happened without the intervention?

Experimental options (or research designs):

Control Group: comparing an untreated research sample against all other groups or samples in the research.

Quasi-experimental options (or research designs):

Difference in Difference (or Double Difference): the before-and-after difference for the group receiving the intervention (where they have not been randomly assigned) is compared to the before-after difference for those who did not.

Instrumental Variables: a method used to estimate the causal effect of an intervention.

Judgemental Matching: a comparison group is created by finding a match for each person or site in the treatment group based on researcher judgements about what variables are important.

Matched Comparisons: participants are each matched with a non-participant on variables that are thought to be relevant. It can be difficult to adequately match on all relevant criteria.

Propensity Scores: statistically creating comparable groups based on an analysis of the factors that influenced people's propensity to participate in the program.

Sequential Allocation: a treatment group and a comparison group are created by sequential allocation (e.g. every 3rd person on the list).

Statistically Created Counterfactual: developing a statistical model, such as a regression analysis, to estimate what would have happened in the absence of an intervention.

Regression Discontinuity: comparing the outcomes of individuals just below the cut-off point with those just above the cut-off point.

Non-experimental options:

Key Informant: asking experts in these types of programmes or in the community to predict what would have happened in the absence of the intervention.

Logically constructed counterfactual: using the baseline as an estimate of the counterfactual. Process tracing can support this analysis at each step of the theory of change.

Approaches: Randomized Controlled Trial (RCT): creating a control group and comparing this to one or more treatment groups to produce an unbiased estimate of the net effect of the intervention.

3. Investigate possible alternative explanations

How will you investigate alternative explanations?

Force Field Analysis: providing a detailed overview of the variety of forces that may be acting on an organizational change issue.

General Elimination Methodology: this involves identifying alternative explanations and then systematically investigating them to see if they can be ruled out.

Key Informant: asking experts in these types of programmes or in the community to identify other possible explanations and/or to assess whether these explanations can be ruled out.

Process Tracing: ruling out alternative explanatory variables at each step of the theory of change.

Ruling Out Technical Explanations: identifying and investigating possible ways that the results might reflect technical limitations rather than actual causal relationships.

Searching for Disconfirming Evidence/Following Up Exceptions: Treating data that don't fit the expected pattern not as outliers but as potential clues to other causal factors and seeking to explain them.

Statistically Controlling for Extraneous Variables: where an external factor is likely to affect the final outcome, it needs to be taken into account when looking for congruence.

Approaches: these approaches combine ruling out possible alternative explanations with options to check the results support causal attribution.

Contribution Analysis: assessing whether the program is based on a plausible theory of change, whether it was implemented as intended, whether the anticipated chain of results occurred and the extent to which other factors influenced the program's achievements.

Collaborative Outcomes Reporting: mapping existing data against the theory of change, and then using a combination of expert review and community consultation to check for the credibility of the evidence.

Multiple Lines and Levels of Evidence (MLLE): reviewing a wide range of evidence from different sources to identify consistency with the theory of change and to explain any exceptions.

Rapid Outcomes Assessment: assessing and mapping the contribution of a project's actions on a particular change in policy or the policy environment.

Find options (methods), resources and more information on these tasks and approaches online at http://betterevaluation.org/plan/understandcauses

BetterEvaluation

Synthesize data from one or more evaluations

Combine data to form an overall assessment of the merit or worth of the intervention, or to summarize evidence across several evaluations.

1. Synthesize data from a single evaluation

How will you synthesize data from a single evaluation?

Processes

Consensus Conference: a process where a selected group of lay people (non-experts) representing the community are briefed, consider the evidence and prepare a joint finding and recommendation.

Expert Panel: a process where a selected group of experts consider the evidence and prepare a joint finding.

Techniques

Cost Benefit Analysis: compares costs to benefits, both expressed in monetary units.

Cost-Effectiveness Analysis: compares costs to the outcomes expressed in terms of a standardized unit (eg additional years of schooling).

Cost Utility Analysis: a particular type of cost-effectiveness analysis that expresses benefits in terms of a standard unit such as Quality Adjusted Life Years.

Multi-Criteria Analysis: a systematic process to address multiple criteria and perspectives.

Numeric Weighting: developing numeric scales to rate performance against each evaluation criterion and then add them up for a total score.

Qualitative Weight and Sum: using qualitative ratings (such as symbols) to identify performance in terms of essential, important and unimportant criteria

Rubrics: using a descriptive scale for rating performance that incorporates performance across a number of criteria.

Value for Money: a term used in different ways, including as a synonym for cost-effectiveness, and as systematic approach to considering these issues throughout planning and implementation, not only in evaluation.

Approaches

Social Return on Investment: a systematic way of incorporating social, environmental, economic and other values into decision-making processes

Examples of overall judgement as success or failure:

	Achieved all intended outcomes	Achieved some intended outcomes	Significant negative outcomes	Overall performance
Scenario 1	✓	-	✗	SUCCESS
Scenario 2	✗	✗	✓	FAIL
Scenario 3	✗	✓	✗	?
Scenario 4	✓	-	✓	?

2. Synthesize data across evaluations

Do you need to synthesize data across evaluations? If so, how should this be done?

Best evidence synthesis: a synthesis that, like a realist synthesis, draws on a wide range of evidence (including single case studies) and explores the impact of context, and also builds in an iterative, participatory approach to building and using a knowledge base.

Meta-analysis: a statistical method for combining numeric evidence from experimental (and sometimes quasi-experimental studies) to produce a weighted average effect size.

Meta-ethnography: a method for combining data from qualitative evaluation and research, especially ethnographic data, by translating concepts and metaphors across studies.

Rapid evidence assessment: a process that is faster and less rigorous than a full systematic review but more rigorous than ad hoc searching, it uses a combination of key informant interviews and targeted literature searches to produce a report in a few days or a few weeks.

Realist synthesis: synthesizing all relevant existing research in order to make evidence-based policy recommendations.

Systematic review: a synthesis that takes a systematic approach to searching, assessing, extracting and synthesizing evidence from multiple studies. Meta-analysis, meta-ethnography and realist synthesis are different types of systematic review.

Vote counting: comparing the number of positive studies (studies showing benefit) with the number of negative studies (studies showing harm).

3. Generalize findings

How can the findings from this evaluation be generalized to the future, to other sites and to other programs?

Analytical generalisation: making projections about the likely transferability of findings from an evaluation, based on a theoretical analysis of the factors producing outcomes and the effect of context. Realist evaluation can be particularly important for this.

Statistical generalisation: statistically calculating the likely parameters of a population using data from a random sample of that population.

Approaches

Positive Deviance: Involves intended evaluation users in identifying 'outliers' – those with exceptionally good outcomes - and understanding how they have achieved these.

Horizontal Evaluation: An approach that combines self-assessment by local participants and external review by peers

Find options (methods), resources and more information on these tasks and approaches online at http://betterevaluation.org/plan/synthesize_value

BetterEvaluation

Report and Support use of findings

Develop and present findings in ways that are useful for the intended users of the evaluation, and support them to make use of them.

1. Identify reporting requirements

What timeframe and format is required for reporting?

Communication plan: developing a plan that outlines the strategies which will be used to communicate the results of your evaluation.

Reporting needs analysis: working with your client to determine their reporting needs.

2. Develop reporting media

What types of reporting formats will be appropriate for the intended users?

Written

Executive Summaries: including an executive summary which is a shortened version of the full report.

Final Reports: ensuring they are readable, straight to the point, and use a writing style that promotes understanding regardless who the target audience is.

Interim reports: presenting the interim, preliminary, or initial evaluation findings.

Memos and email: maintaining ongoing communication among evaluation stakeholders through brief and specific messages about a particular issue.

News media communications: sharing news relating to evaluation findings through press releases.

Newsletters, bulletins, briefs and brochures: highlighting particular findings or angles on an evaluation using shorter communications such as bulletins, briefs, newsletters, blogs and brochures.

Postcards: collecting information quickly in order to provide a short report on evaluation findings (or an update on progress).

Website communications: disseminating information such as that coming from evaluations via a range of web based tools.

You may develop a number of reports, in different formats, for different sets of stakeholders. Work with your primary users and stakeholders to determine when and in what form they want to receive evaluation reports. Also determine who you will involve in viewing draft and interim reports.

Presentations

Conference: discussing a set topic or theme in a large group of people at a set venue.

Displays and exhibits: drawing attention to particular issues and assisting in community engagement.

Flip Charts: providing a useful way of interacting with your audience and therefore allowing you to present your own ideas and results and also to immediately record input, feedback and ideas from your audience.

Information Contacts: providing a contact person for all media and public enquiries about a project or program.

Posters: presenting your evaluation findings in the form of a poster provides a good opportunity to get your message across in a clear way while also providing opportunities for feedback.

PowerPoint: organizing and communicate information coming from evaluations in the form of a slide show which can be used at a meeting or conference.

Teleconference: facilitating discussion of evaluation findings via telephone.

Verbal briefings: providing specific information to an audience of interested participants allowing for a structured question and answer format based on that information.

Video: highly flexible and immediate medium which allows you to make an emotional meaningful connection with the audience.

Videoconference: gathering data, communicating information about an evaluation, reporting findings, receiving feedback, and planning for utilization.

Web-conference: bringing people together from around the world using the internet.

Creative

Cartoons: allowing readers to see a point differently, add humour, and break up large sections of prose.

Photographic reporting: making your report more appealing to readers and also making the key messages more memorable by including photographs.

Poetry: communicating the experience of participants can be achieved by presenting some of the findings in the form of a poem.

Reporting in pictures: presenting information in an alternative way and therefore increasing understanding of your results.

Theatre: communicating evaluation findings and engaging intended users in responding to them.

Presenting your report in a creative manner may be the most relevant means to get your information across if the context allows for it. You may consider working with an artist or a graphic designer to produce creative displays.

Graphic design

Arrangement: Descriptive text and its related data visualization should be arranged so they appear together on a page. Narrative text should be left-justified.

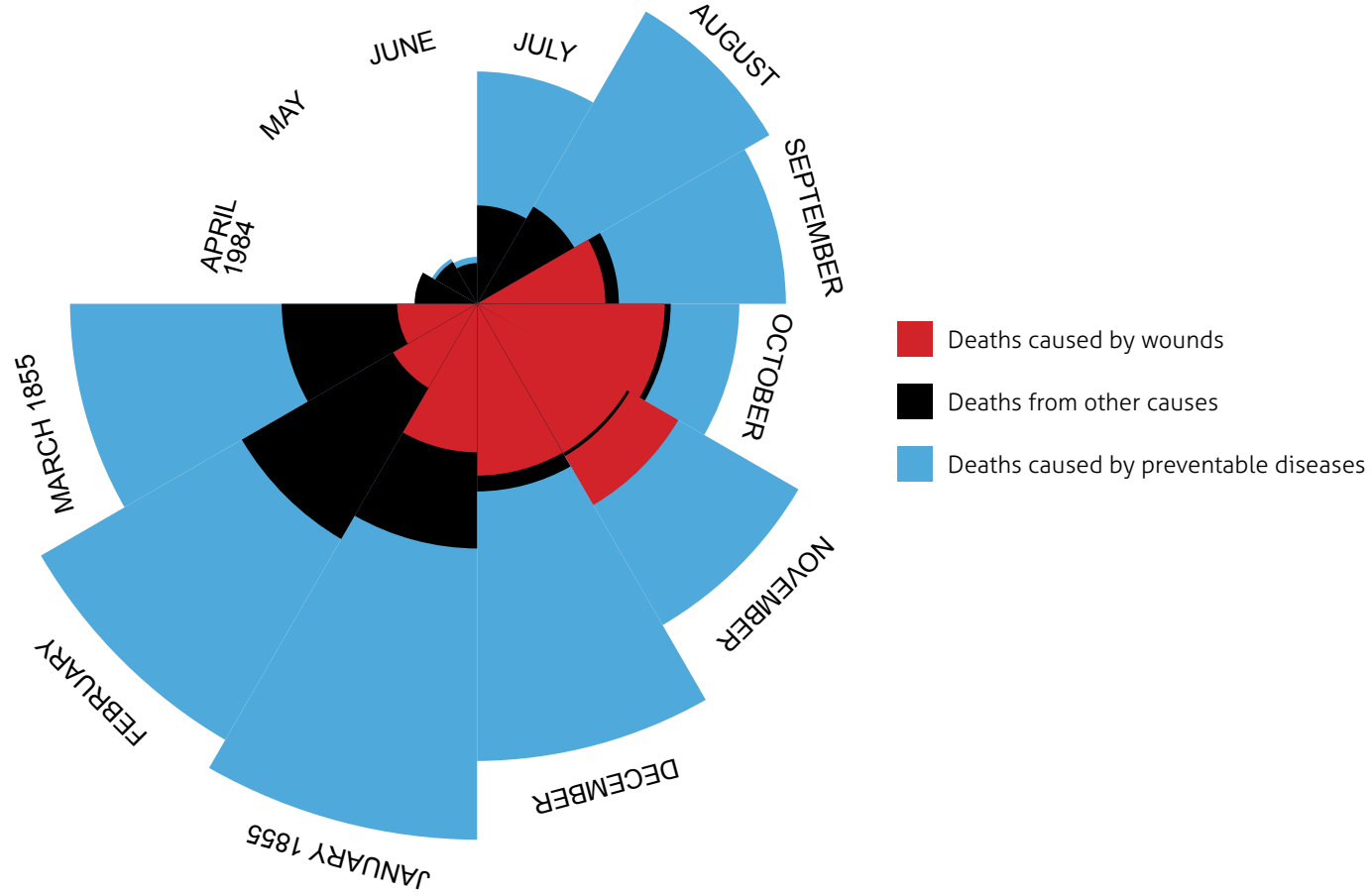
Color: Blocks of background color can help group cognitively-similar items or set off reporting elements like sidebars. Text intended for narrative reading should be set in black or dark gray on a white or very light background.

Images: Written reports and presentations should always include images. Beyond just charts and graphs, photographs or drawings increase the relevancy of the material to the audience and make the report more engaging.

Type: Generally speaking, serif fonts support readability in long, narrative-style documents produced on paper. Sans serif fonts are easier to read in electronic reporting media.

Example of an infographic used as a reporting format

This graphic was developed in the 1850s by Florence Nightingale depicting causes of death in the British Army in 1854. The graph shows that by far the biggest killer was preventable disease, not battle wounds as was previously thought. This led to improved conditions in military hospitals.



3. Ensure accessibility

How can the report be easy to access and use for different users?

Applied graphic design principles

Simplified report layout: three different ways of simplifying the report layout are to eliminate chartjunk, emphasise headings as summary statements, and use descriptive subtitles.

One-Three-Twenty-Five (1:3:25) principle: ensuring that research findings are presented in a logical and consistent manner by allowing for a 1 page outline, a 3 page executive summary and 25 pages to present the findings and methodology.

Support users with auditory disabilities

Support users with colour blindness

Support users with visual disabilities

Use appropriate language: ensuring the language of a report is clear, concise and allows accessibility for all stakeholders.

4. Develop recommendations

Will the evaluation include recommendations? How will these be developed and by whom?

Beneficiary exchange: discussing findings between beneficiaries in order to provide feedback.

Chat rooms: setting up online spaces where findings can be discussed.

Electronic democracy: using new and emergent forms of media in order to engage community members in seeking to influence the decision making process.

External review: having external experts or anonymous reviewers provide feedback.

Group critical reflection: facilitating a group stakeholder feedback session.

Individual critical reflection: asking particular individual stakeholders for their independent feedback.

Lessons learned

Participatory recommendation screening: testing recommendations with key stakeholders.

World Cafe: hosting a group dialogue which emphasizes the power of simple conversation when considering relevant questions and themes.

5. Support use

In addition to engaging intended users in the evaluation pocess, how will you support the use of evaluation findings?

Annual reviews: reviewing major evaluation findings and conclusions based on evaluation studies completed during the preceding year.

Conference Co-presentations: evaluators and evaluation commisioners or users jointly presenting findings or discussions about processes from an evaluation.

Policy briefings: providing evaluation findings and lessons learned in an accessible manner for target audiences which can be followed up by management and staff.

Recommendations tracking: keeping a transparent record of the responses to and action from recommendations.

Social Learning: focusing on how people learn through social interactions, such as modelling, making connections, sharing experiences and resources, collaboration and self-organization.

Trade Publications: producing a non-technical version of key findigns for a publication aimed at staff who can use the findings.

Find options (methods), resources and more information on these tasks and approaches online at http://betterevaluation.org/plan/reportandsupportuse