unicef
Office of Research-Innocenti

Methodological Briefs
Impact Evaluation No. 6

Overview: Strategies for Causal Attribution

Patricia Rogers

RMIT UNIVERSITY
3ie International Initiative for Impact Evaluation
BetterEvaluation
unicef Office of Research-Innocenti

UNICEF OFFICE OF RESEARCH

The Office of Research is UNICEF's dedicated research arm. Its prime objectives are to improve international understanding of issues relating to children's rights and to help facilitate full implementation of the Convention on the Rights of the Child across the world. The Office of Research aims to set out a comprehensive framework for research and knowledge within the organization, in support of UNICEF's global programmes and policies, and works with partners to make policies for children evidence-based. Publications produced by the Office are contributions to a global debate on children and child rights issues and include a wide range of opinions.

The views expressed are those of the authors and/or editors and are published in order to stimulate further dialogue on impact evaluation methods. They do not necessarily reflect the policies or views of UNICEF.

OFFICE OF RESEARCH METHODOLOGICAL BRIEFS

UNICEF Office of Research Methodological Briefs are intended to share contemporary research practice, methods, designs, and recommendations from renowned researchers and evaluators. The primary audience is UNICEF staff who conduct, commission or interpret research and evaluation findings to make decisions about programming, policy and advocacy.

This brief has undergone an internal peer review.

The text has not been edited to official publication standards and UNICEF accepts no responsibility for errors.

To consult and download the Methodological Briefs, please visit http://www.unicef-irc.org/KM/IE/

For readers wishing to cite this document we suggest the following form:
Rogers, P. (2014). Overview: Strategies for Causal Attribution, *Methodological Briefs: Impact Evaluation 6*, UNICEF Office of Research, Florence.

Acknowledgements: This brief benefited from the guidance of many individuals. The author and the Office of Research wish to thank everyone who contributed and in particular the following:

Contributors: Simon Hearn, Greet Peersman, Jessica Sinclair Taylor, Howard White
Reviewers: Nikola Balvin, Sudhanshu Handa, Debra Jackson, David Parker

September 2014

UNICEF Office of Research - Innocenti
Piazza SS. Annunziata, 12
50122 Florence, Italy
Tel: (+39) 055 20 330
Fax: (+39) 055 2033 220
florence@unicef.org
www.unicef-irc.org

1. CAUSAL ATTRIBUTION: A BRIEF DESCRIPTION

One of the essential elements of an impact evaluation is that it not only measures or describes changes that have occurred but also seeks to understand the role of particular interventions (i.e., programmes or policies) in producing these changes. This process is often referred to as causal attribution, causal contribution or causal inference. This brief provides an overview of different ways to examine causal attribution, using a combination of research design and particular data collection and analysis strategies.

The OECD-DAC definition of impact makes it clear that an impact evaluation must establish what has been the cause of observed changes: "Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended."[1]

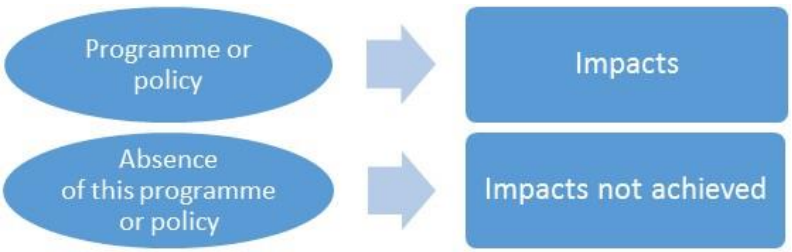
Causal attribution is defined by OECD-DAC as: "Ascription of a causal link between observed (or expected to be observed) changes and a specific intervention."[2]

This definition does not require that changes are produced solely or wholly by the programme or policy under investigation[3]. In other words, it takes into consideration that other causes may also have been involved, for example, other programmes/policies in the area of interest or certain contextual factors (often referred to as 'external factors').

Evaluations produce stronger and more useful findings if they not only investigate the links between activities and impacts but also investigate links along the causal chain between activities, outputs, intermediate outcomes and impacts. A 'theory of change'[4] (see Brief No. 2, Theory of Change) is therefore helpful in guiding causal attribution in an impact evaluation. The evaluation may confirm the theory of change or it may suggest refinements based on the analysis of evidence.

It can be useful to bear in mind the three conceptualizations of cause and effect below when planning an impact evaluation.

1. **Sole causal attribution**: This is where the programme or policy is both necessary and sufficient to produce the impacts, independently (or relatively independently) of contextual factors or other interventions.



   In practice, programmes or policies are rarely sufficient to produce the intended impacts alone – and there are often alternative ways to achieve the intended impacts. As a result, this is rarely a useful model of cause and effect for impact evaluation.

---

1 Organisation for Economic Co-operation and Development – Development Assistance Committee, *Glossary of Key Terms in Evaluation and Results Based Management*, OECD-DAC, Paris, 2010. See http://www.oecd.org/development/peer-reviews/2754804.pdf.

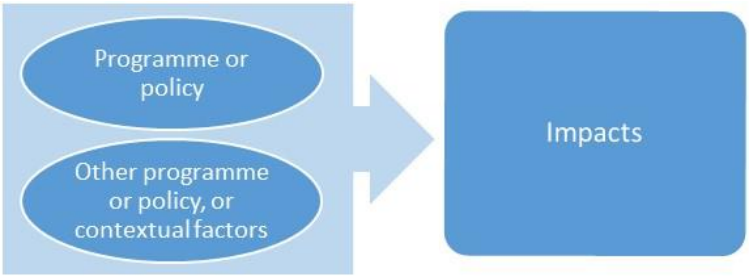
2 Ibid.

3 United Nations Evaluation Group, *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*, Guidance Document, UNEG, New York, 2013. See http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=1434.

4 A theory of change explains how activities are understood to produce a series of results that contribute to achieving the ultimate intended impacts.

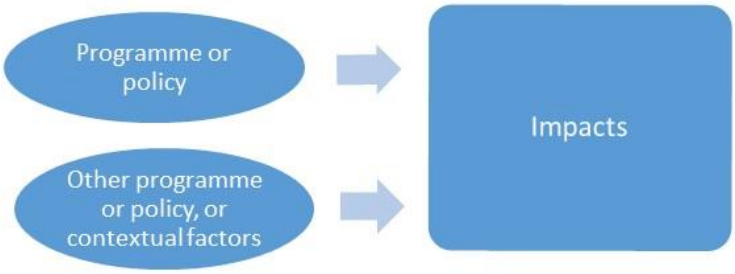
2. **Joint causal attribution**: This is where the programme or policy produces the impacts in conjunction with other programmes or policies, or certain contextual factors. This can include complementary programmes that either lay the foundations for or reinforce the programme that is being evaluated. It can also include factors to do with the implementation environment such as the skills and infrastructure of implementing partners, or to do with the participants, for example, in terms of their levels of motivation and prior knowledge. Where these contributing factors are either absent or negative, the impacts will not be achieved, or will be achieved to a much lesser extent.

   This is a very common situation, where a programme will only work effectively if favourable conditions are present and/or unfavourable conditions are removed. It has important implications for how impact evaluation is conducted and how the findings are used. For example, if a programme is found only to work in a place where there is transparency about government funding allocations, then it should only be implemented in situations where this already exists, or in conjunction with efforts to achieve transparency. In order to identify these other contributing factors, it is important both to develop a good theory of change that includes them – based on previous research and existing knowledge – and to gather data about them.



3. **Alternative (or multiple) causal paths**: This is where the programme or policy is just one of several possible ways to achieve the impacts. A particular programme or policy can produce the impacts, but they can also come about through other interventions and/or external factors.

   This can happen when participants are able to access services through an alternative provider, or when different programmes by different providers are intended to produce the same impact. For example, a number of programmes might aim to reduce infant mortality – some by improving nutrition, some by reducing waterborne diseases and some through immunization.



This scenario has important implications for impact evaluation. If a counterfactual design is used, where participants are compared to non-participants, it is important to investigate what services the non-participants are accessing. If a programme is intended to produce impacts that are also the goal of other programmes, it is especially important to identify intermediate outcomes in the theory of change, and to collect data about these. This will help to establish the extent to which impacts have been due to the programme being evaluated.

There are three broad strategies for causal attribution in impact evaluations:

- estimating the counterfactual (i.e., what would have happened in the absence of the intervention, compared to the observed situation)
- checking the consistency of evidence for the causal relationships made explicit in the theory of change
- ruling out alternative explanations, through a logical, evidence-based process.

Using a combination of these strategies can usually help to increase the strength of the conclusions that are drawn (for more on this, see section 3).

It is important that all impact evaluations systematically undertake causal attribution. But there must also be recognition of what can be realistically achieved by a single evaluation – especially when there is limited time for iterative data collection and analysis. A single impact evaluation can add to the evidence base, but even where there is a reasonable degree of confidence that the programme or policy caused the observed impacts, this does not mean that the evaluation has examined all aspects of the intervention or how it will work in other settings or at other times. A synthesis of findings across impact evaluations thus provides stronger evidence than can a single evaluation about the generalizability of the effects observed.

Main points

1. Causal attribution investigates the causal links between a programme or other intervention and observed changes.
2. Causal attribution is an essential element of impact evaluation.
3. There are several strategies for examining causal attribution, all of which benefit from being based on a sound theory of change.
4. The 'best fit' strategy for causal attribution depends on the evaluation context as well as what is being evaluated.

2. WHEN IS CAUSAL ATTRIBUTION APPLIED IN IMPACT EVALUATION?

Causal attribution is an essential element of any impact evaluation. It enables an evaluation to report not only that a change occurred, but also that it was due, at least in part, to the programme or policy being evaluated.

A change cannot be claimed to be an impact unless there is a demonstrated link between it and the intervention. This is aided by demonstrated connections between the intervention's inputs and outputs and the outcomes and impacts that are observed (see Brief No. 1, Overview of Impact Evaluation).

If an impact evaluation fails to systematically undertake causal attribution, there is a greater risk that the evaluation will produce incorrect findings and lead to incorrect decisions – for example, deciding to scale up when the programme is actually ineffective or effective only in certain limited situations, or deciding to exit when a programme could be made to work if limiting factors were addressed.

3. HOW TO UNDERTAKE CAUSAL ATTRIBUTION

In all cases, it is useful to begin by identifying or revisiting the pathway of change by which the *intervention* (programme or policy) is understood to contribute to the intended or observed impacts.

The pathway of change can be described in a causal chain – or programme logic model – and it can be helpful to frame this using the standard results-based management (RBM) categories of inputs, outputs, outcomes and impact. The evaluation can elucidate the pathway of change where one has not been explicitly phrased, or it can validate (or disprove) a hypothesized pathway, which is often designated as a theory of change (see Brief No. 2, Theory of Change).

Several different strategies may be used to undertake causal attribution, each of which has its own particular strengths, limitations and suitability according to the specific programme and evaluation context. While there are different ways of classifying designs and methods for causal attribution in impact evaluation[5], for the purposes of providing an overview in this brief these are discussed in terms of three broad approaches:

- **counterfactual approaches** – these develop an estimate of what would have happened in the absence of a programme or policy, and compare this to what has been observed in the presence of the intervention. This approach requires the use of a control group or comparison group.
- **consistency of evidence with causal relationship** – this identifies patterns that would be consistent with a causal relationship, which is usually grounded in a well developed theory of change, and then seeks confirming and disconfirming evidence.
- **ruling out alternatives** – this identifies possible alternative causal explanations and seeks information to see if these can be ruled out.

These approaches are discussed in more detail in this brief and other briefs in this series (see Brief No. 7, Randomized Controlled Trials; Brief No. 8, Quasi-experimental Design and Methods; and Brief No. 9, Comparative Case Studies).

Counterfactual approaches

A counterfactual approach involves developing an estimate of what would have happened in the absence of a programme or policy, and comparing this to what has been observed in the presence of the intervention. Four types of evaluation design apply a counterfactual approach and these vary according to how the 'counterfactual' (non-intervention) effect is estimated.

Experimental designs

These involve the randomized assignment of participants to test the effects of an intervention. Participants are randomly allocated either to one or more groups participating in the intervention, or to a control group that receives no intervention. Such evaluations are called randomized controlled trials (RCTs; see also Brief No. 7, Randomized Controlled Trials) because of the rigorous manner in which the intervention is applied – similar to how pharmaceutical products are tested.

RCTs reduce the risk of selection bias, where participants and non-participants may have systematic differences in addition to whether they receive the intervention or not – and thus where the apparent impact of the programme or policy may be due to these differences rather than (or in addition) to its actual effect.

[5] See, for instance, the framework outlined by Elliott Stern et al., 'Broadening the range of designs and methods for impact evaluations, *DFID Working Paper* No. 38, Department for International Development, London, 2012. This working paper distinguishes between experimental, statistical, theory-based, case-based, participatory and synthesis studies.

Quasi-experimental[6] designs

These involve constructing comparison groups in various ways that do not involve random assignment (see also Brief No. 8, Quasi-experimental Design and Methods). These designs are often more feasible in an impact evaluation and may be considered to provide a sufficiently valid comparison between those who do and those who do not receive an intervention. Options for constructing comparison groups include judgemental matching[7] and matched comparisons[8] as well as other, more commonly used options:

- **judgemental matching** – creating a comparison group by finding a match for each person or site in the treatment group based on researcher judgements about what variables are important.
- **matched comparisons** – matching participants (individuals, organizations or communities) with a non-participant on variables that are thought to be relevant.
- **propensity score matching (PSM)** – statistically creating comparable groups based on an analysis of the factors that influenced people's propensity to participate in the intervention.
- **regression discontinuity design** – comparing the outcomes of those individuals just below and just above a cut-off point.
- **sequential allocation** – a treatment group and a comparison group are created by sequential allocation (e.g., every third person on the list).

Options for analysis in quasi-experimental designs include:

- **single difference** – comparing the outcomes in the treatment group with the outcomes in the comparison group at a single point in time following the intervention
- **double difference (also known as difference-in-differences or DID)** – comparing the 'before and after' difference for the group receiving the intervention (where they have not been randomly assigned) to the before and after difference for those who did not.

Hypothetical counterfactual

Sometimes it is possible to construct a 'hypothetical counterfactual' of what would have happened in the absence of the programme or policy by demonstrating that conditions would have remained the same.

For example, in an impact evaluation of a water project that included studying its impact on time spent carrying water, it was sufficient to demonstrate that the time spent carrying water had been reduced after a central pump had been installed as it was reasonable to assume that in the absence of the pump this time would not have changed. In practical terms, this represents a before and after comparison (which may be among the same population) rather than a comparison of different groups at a given point in time.

Modelling

A statistically created counterfactual entails developing a statistical model such as a regression analysis to estimate what would have happened in the absence of an intervention (see Brief No. 13, Modelling).

---

6 Some literature refers to these designs as 'non-experimental' rather than 'quasi-experimental'. The author of this brief prefers to use the term quasi-experimental since the designs in many ways resemble experiments, and because non-experimental approaches to causal attribution are quite different and are therefore usefully categorized separately.

7 Creating a comparison group by finding a match for each person or site in the treatment group based on researcher judgements about what variables are important.

8 Matching each participant (individual, organization or community) with a non-participant on variables thought to be relevant.

Consistency of evidence with causal relationship

This approach to causal attribution involves identifying what evidence would be consistent with a causal relationship, and then gathering and analysing data from different sources to determine whether the evidence matches this. This approach is normally guided by a theory of change, whether this is elaborated on in detail or implicit in the programme or policy logic model.

Different data collection and analysis methods can be used to assemble this evidence. It is advisable to combine several of these methods in a single impact evaluation according to the level of certainty required and the possible counter-explanations identified. In addition, evidence from previous research and evaluation can be used – for example, impact evaluations of vaccination programmes do not need to test every link in the causal chain given the body of knowledge available from previous research.

Options include:

- **achievement of intermediate outcomes** – checking whether all cases that achieved the final impacts also achieved the intermediate outcomes identified in the theory of change.
- **checking results against expert predictions** – making predictions based on the theory of change or an emerging theory of wider contributors to outcomes, and then following up on whether or not these predictions actually materialize over time.
- **checking timing of impacts** – determining whether the timing of impacts is consistent with a causal relationship, again with reference to the theory of change. For example, the impact occurs a reasonable time after the programme or policy has taken effect.
- **comparative case studies** – systematically comparing case studies to understand the array of factors that may be responsible for the impacts.
- **dose-response patterns** – examining the link between the 'dose' (the intensity or level at which the intervention has been applied) and the 'response' (the effect observed) as part of efforts to determine whether or not the intervention caused the outcome. This is another method drawn from the world of biological testing, which is increasingly applied in social science research and evaluation.
- **checking consistency with existing literature** – checking results against what is known from reviewing the literature in the area of interest, to identify consistencies/inconsistencies. This must be done cautiously and with explicit mention of any limitations in the existing literature (or subset thereof).
- **interviewing key informants** – this is not about asking key informants if they believe the intervention has produced the impacts (which can be affected by their level of knowledge of causal processes and their intentions regarding the continuation of the intervention). It is instead about asking them to explain the causal processes following their involvement. For example, in a programme that aims to support the national government to develop new, evidence-based policies, a key informant interview might ask about the process of developing the policies, and work backwards to the programme, avoiding leading questions that assume the programme has been the major cause. These interviews can provide evidence to explain, for example, how a training programme made a difference in terms of the government's subsequent ability to develop and negotiate particular policies.
- **modus operandi** – drawing on the previous experience of participants and stakeholders to determine what array or pattern of effects is typical for an intervention. For example, a training programme might have a particular 'signature' terminology or practice that is evident among participants.
- **process tracing** – developing alternative hypotheses and then gathering evidence (clues) to determine whether or not these are compatible with the available hypotheses. (Sherlock Holmes's approach to detective work exemplifies this logic.)

- **qualitative comparative analysis** – related to comparative case studies, this involves comparing the configurations of different case studies to identify the components that appear to be most responsible for producing specific outcomes.
- **realist analysis of testable hypotheses** – using a realist theory of change (i.e., what works for whom in what circumstances and through what causal mechanisms) to identify specific contexts in which positive results would and would not be expected, and checking these against the observed situations.

For more information, see Brief No. 9, Comparative Case Studies and Brief No. 13, Modelling. For more information on the other methods outlined above, see the BetterEvaluation web page 'Check the results support causal attribution'.

Ruling out alternative explanations

The third strategy for causal attribution is to identify possible alternative explanations for the achievement of impacts and then gather data to see if these can be ruled out. This strategy is particularly useful to apply when the available evidence may be sufficient only to suggest 'correlation' but not 'causality'.

Options include:

- **key informant interviews** – asking either experts in the specific type of programme, community members or other stakeholders to identify other possible explanations and, if feasible, to assess whether these explanations can be ruled out.
- **process tracing** – using evidence to rule out alternative explanatory variables at each step of the theory of change.
- **ruling out technical explanations** – identifying and investigating possible ways that the observed results might reflect technical limitations of the data (for example, regression to the mean, or unreliable measurement) rather than causal relationships.
- **modelling** – investigating alternative explanations through means of statistical analysis such as regression or logistic regression to control for confounding factors.
- **general elimination methodology** – this is carried out in two stages, (1) identifying possible explanations (including that the observed changes are indeed due to the intervention, plus as many alternative explanations as possible) using a *combination* of options such as those listed above (e.g., key informant interviews *and* brainstorming, *and* reviewing previous evaluations/research); and (2) gathering and analysing data to see if the possible alternative explanations can be ruled out. This option for causal attribution is most effective when used in combination with the other strategies.

As the use of systematic, non-experimental causal attribution in development impact evaluations is rare, this approach is best illustrated by an example from another area of evaluation. An impact evaluation of new compulsory cycle helmet legislation[9] established that the number of cyclist head injuries had declined following the introduction of the law. This was not in itself, however, evidence of a causal relationship. The evaluation could not draw on a counterfactual approach, since there was no suitable comparison group. It could draw partly on the consistency of evidence with a causal relationship – i.e., confirming that more helmets had been bought since the law came into effect, and that observations showed a high level of compliance with the new legislation.

The main strategy for causal attribution in this case, however, was to identify and investigate possible alternative explanations, with the aim of ruling these out. For example, the reduction in cyclist head injuries might have been achieved through a reduction in the rate of cycling, as people decided that they would

[9] Walter, Scott R., et al., 'The impact of compulsory cycle helmet legislation on cyclist helmet head injuries in New South Wales, Australia', *Accident Analysis and Prevention*, 43, 2011, pp. 2064–2071.

rather not cycle if they had to wear a helmet. It was not feasible to undertake the large survey needed to gather new data to investigate this alternative explanation. Existing data were available that were relevant and sufficiently accurate, however. If the decline in head injuries had been caused by a reduced level of cycling, the total number of cycling injuries would also have been expected to have declined. Instead the data showed that only the number of head injuries had declined, strengthening the argument that the reduction had not been brought about by reducing the rate of cycling.

4. HOW TO CHOOSE THE BEST STRATEGY FOR CAUSAL ATTRIBUTION

There are differing views about the relative merits of the different strategies and options for causal attribution, which have implications for the choices made when embarking on joint evaluations with other agencies. Some organizations rank these approaches in order of merit, requiring a counterfactual approach, and labelling experimental designs involving a control group as the 'gold standard'.

For example, the USAID Evaluation Policy clearly requires a constructed counterfactual group, and the use of a control group (i.e., RCT) is preferred for an evaluation to be considered an impact evaluation:

> "Impact evaluations measure the change in a development outcome that is attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors."10

Other organizations may agree on the technical merits of different strategies but allow for the best choice to be situationally responsive to the nature of what is being evaluated and the circumstances of the evaluation.

For example, the discussion paper on impact evaluation by AusAID (now the Department of Foreign Affairs and Trade, Australia) proposes using a combination of methods and designs that are situationally appropriate:

> "AusAID recognises that **a range of methodological approaches can be used for impact evaluation.** The need for an impact evaluations starts with a need for evidence (as detailed above). The purpose and questions of the evaluation, the complexity of the intervention and the context of the intervention, will determine what types of methods are used. Impact evaluations in AusAID should not be methods-driven."[11]

The recent UNEG Impact Evaluation Guidance Document advocates choosing a mix of methods and designs appropriate to the circumstances of the impact evaluation and the types of questions it seeks to answer:

> "When undertaking impact evaluation, even more so than in other types of evaluation, it is important to do more than list a set of methods and assume that this amounts to a methodology. Rather, within the chosen evaluation design, there should be an explicit over-arching methodological framework, which enables the individual methods to be brought together to produce a meaningful overall analysis that can evaluate whether the intervention has had impacts, or made a contribution towards them. It is essential to tailor the particular evaluation design and mix of

[10] U.S. Agency for International Development, Evaluation, Learning from Experience, Evaluation Policy, USAID, Washington, D.C., 2011. See http://www.usaid.gov/sites/default/files/documents/1868/USAIDEvaluationPolicy.pdf.

[11] AusAID Office of Development Effectiveness, 'Impact Evaluation: A Discussion Paper for AusAID Practitioners', Discussion Paper, AusAID ODE, 2012. See http://www.ode.dfat.gov.au/publications/impact-evaluation.html.

> methods, to the particular situation, context, questions and issues of most concern."[12]

UNICEF follows UNEG guidance, and so the choice of which strategy or strategies to use for causal attribution should therefore be based on an assessment of what is most appropriate for the situation. This assessment must consider the nature of the:

- programme or policy being evaluated – over what populations or areas is the programme or policy being applied? What are the units of analysis? How variable is the programme's implementation?
- evaluation – what stage is the programme or policy at? Have participants already engaged with the programme or policy? What data are already available? What resources are available in terms of time, money for external evaluators, and skills? Do other organizations involved in the evaluation have particular requirements for the impact evaluation design?

It is also possible to combine the three basic strategies in a single evaluation. Contribution analysis[13] is an overall approach that can encompass different strategies. It works through a structured process of documenting and bringing together what is already known in terms of the current theory of change, evidence from previous research and evaluations, and existing data from a programme or policy, assesses the quality of the contribution story, and then gathers evidence systematically to address identified gaps and weaknesses.

5. EXAMPLE OF GOOD PRACTICES

The evaluation of the Paris Declaration on Aid Effectiveness[14] provides an example of systematic causal attribution in a situation where it was not possible to identify or construct a credible counterfactual. The aim was to evaluate whether or not the Paris Declaration had been successfully implemented, and the extent to which, how and why it had contributed to development results.

The theory of change described how programmatic actions would lead to intermediate outcomes (in improved aid effectiveness) and longer-term impacts (in contributions to improved development results) and involved several complex pathways designed to seek out important potential contextual factors that might drive or inhibit change.

Emphasis was placed on the structured way in which the evaluation teams were to use a mixed methods approach to assess "plausible contributions" made by the Paris Declaration to development results in each context, and on providing "clear evidence of any changes and connections observed and any other plausible explanations"[15].

A comprehensive evaluation framework set out the types of evidence that evaluators should look for and the methods or forms of analysis that could be applied. It also included a rating system to indicate the relevance of the evidence found to key evaluation questions, the extent to which it could be triangulated and therefore considered reliable, and the degree to which data were from recent, credible sources, and to indicate the extent to which data collection methods and analysis provided a reasonable basis for the findings and conclusions drawn.

---

12 United Nations Evaluation Group, *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*, Guidance Document, UNEG, New York, 2013. See http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=1434.

13 Mayne, John, 'Contribution analysis: An approach to exploring cause and effect', ILAC Brief 16, Institutional Learning and Change, 2008. See http://www.cgiar-ilac.org/files/ILAC_Brief16_Contribution_Analysis_0.pdf.

14 Wood, Bernard, et al., *The Evaluation of the Paris Declaration*, Final Report, Danish Institute for International Studies, Copenhagen, May 2011. See http://www.oecd.org/derec/dacnetwork/48152078.pdf.

15 White, Howard and Daniel Phillips, 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework', *International Initiative for Impact Evaluation Working Paper* No. 15, 3ie, New Delhi, 2012, p. 12. See http://www.3ieimpact.org/media/filer/2012/06/29/working_paper_15.pdf.

Methods employed included literature and document reviews, quantitative/statistical analysis of the most relevant available data, survey instruments, interviews, focus groups and stakeholder analysis.

6. EXAMPLES OF CHALLENGES

UNICEF impact evaluations often display inadequacies in terms of causal attribution. The following examples are all from anonymized, completed UNICEF impact evaluations.

Failure to systematically address causal attribution

Some impact evaluations make no attempt to examine causal attribution but still imply or claim attribution. Other impact evaluations do not address causal attribution if it is not possible to create a comparison or control group. This is not only a lost opportunity – given that there are several alternative ways to address causal attribution, as discussed above – but it also defeats the purpose of undertaking an impact evaluation. If there is no attempt to address causal attribution, or it is done inappropriately, this may lead to spurious conclusions being made about the effectiveness of the programme or policy and misguided decision making.

For example, one evaluation acknowledged that it had not investigated causal relationships, but claimed that the programme had been effective since the intended change (increased service usage) had been achieved.

Failure to accurately describe the type of counterfactual used

Some impact evaluations use the term 'control group' to refer to any comparison group – including those that are poor matches for the treatment or intervention group – and provide insufficient information to allow the quality of the matching to be judged. For example, one evaluation referred to "control sites"[16] and acknowledged that these had been constructed using quasi-experimental techniques rather than random assignment, but then failed to provide any information about how they had been selected or constructed or if their comparability to the "experimental sites"[17] had been checked.

This is problematic because control and comparison groups are constructed in very different ways (as explained above). This has implications for how well each constructed group might match the intervention group at baseline (see Brief No. 7, Randomized Controlled Trials and Brief No. 8, Quasi-experimental Design and Methods) and therefore accurately represent the real impact of a programme or policy.

Failure to seek out or try to explain evidence that does not fit the theory of change

Some impact evaluations do not seek out or try to explain evidence that is inconsistent with a causal relationship. For example, one evaluation claimed that a capacity-building programme had produced certain impacts within an organization even though it reported that the programme officer had spent little time with the organization and had not provided any assistance to the staff or management, and that another programme had provided the organization with training that was seen to have developed its capacity.

16 No source provided as impact evaluations have been anonymized.

17 Ibid.

7. KEY READINGS AND LINKS

Other particularly relevant briefs in this series:

- Randomized Controlled Trials (Brief No. 7)
- Quasi-experimental Design and Methods (Brief No. 8)
- Comparative Case Studies (Brief No. 9)
- Modelling (Brief No. 13).

BetterEvaluation, 'Understand Causes', web page, BetterEvaluation, http://betterevaluation.org/plan/understandcauses.

European Commission, *Evalsed Sourcebook: Methods and Techniques*, European Commission, 2013. See http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf.

Gaarder, Marie, and Jeannie Annan, 'Impact Evaluation of Conflict Prevention and Peacebuilding Interventions', *Policy Research Working Paper* No. 6496, Public Sector Evaluation Group, Independent Evaluation Group, World Bank Group, 2013. See http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-6496.

Hughes, Karl, and Claire Hutchings, 'Can we obtain the required rigour without randomisation? Oxfam GB's non-experimental Global Performance Framework', *International Initiative for Impact Evaluation Working Paper* No. 13, 3ie, New Delhi, 2011. See http://www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_13.pdf.

Mayne, John, 'Contribution analysis: An approach to exploring cause and effect', ILAC Brief 16, Institutional Learning and Change, 2008. See http://www.cgiar-ilac.org/files/ILAC_Brief16_Contribution_Analysis_0.pdf.

Rogers, Patricia J., RMIT University and BetterEvaluation, 'Introduction to Impact Evaluation', Impact Evaluation Notes No. 1, InterAction, 2012. See http://www.interaction.org/document/introduction-impact-evaluation.

Rogers, Patricia J., 'Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation', in Chambers, Robert, et al., 'Designing impact evaluations: different perspectives', *International Initiative for Impact Evaluation Working Paper* No. 4, 3ie, New Delhi, 2009. See http://www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_4.pdf.

Stern, Elliott, et al., 'Broadening the range of designs and methods for impact evaluations', *DFID Working Paper* No. 38, Department for International Development, London, 2012.

White, Howard, and Daniel Phillips, 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework', *International Initiative for Impact Evaluation Working Paper* No. 15, 3ie, New Delhi, 2012, p. 43. See http://www.3ieimpact.org/media/filer/2012/06/29/working_paper_15.pdf.

White, Howard, 'Some Reflections on Current Debates in Impact Evaluation', *International Initiative for Impact Evaluation Working Paper* No. 1, 3ie, New Delhi, 2009. See http://www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_1.pdf.

.

GLOSSARY

Activity	*Actions taken or processes through which inputs are mobilized to produce specific outputs. For example, counselling sessions that adhere to quality standards.*
Brainstorming	*A strategy by which one generates ideas through an intensive, freewheeling discussion. The point is to generate as many ideas as possible – including ones that may initially seem absurd – and analyse and evaluate them at a later stage.*
Case study	*An in-depth examination of a single or more case(s) (e.g. individuals, groups, institutions, countries, processes), conducted and designed to result in a thorough and well-organized understanding of the subject(s) being examined. Case studies can address the micro-situation of a single person in everyday life or the macro-situation of a state and/or even global processes. The results can be used as stand-alone findings, or they can be integrated as inputs into broader syntheses and/or comparative analyses.*
Causal relationship	*The relationship established that shows that an independent variable causes a change in a dependent variable. Establishes, also, how much of a change is shown in the dependent variable.*
Causality/causation	*The principle that one variable (X) produces change in another variable (Y). It is based on the assumption that events occur in a predictable, non-random way, and that one event leads to, or causes, another. To establish causation, the two variables must be associated or correlated with each other; the first variable (X) must precede the second variable (Y) in time and space; and alternative, non-causal explanations for the relationship (such as spurious ones) must be eliminated. Events in the physical and social worlds are generally too complex to be explained by any single factor. For this reason, scientists are guided by the principle of multiple causation, which states that one event occurs as a result of several factors operating or occurring in combination.*
Comparison group	*In a quasi-experimental research design, this is the group of research participants that, for the sake of comparison, does not receive the treatment or intervention given to the intervention group. Comparison group subjects are typically not randomly assigned to their condition, as would be true of control group subjects in an experimental design study. See: control group, treatment group.*
Control group	*Participants in a research study/evaluation who do not receive the experimental treatment/intervention.*
Correlation	*Mutual relationship or association of two or more concepts or variables, such that when one changes in value, the other one does also. Variables may be correlated positively (i.e., they change in the same direction) or negatively (that is, they change in the opposite direction). Correlation is necessary but not sufficient to demonstrate causation.*
Counterfactual approach	*Developing an estimate of what would have happened in the absence of the programme or policy; this implies the use of a control group or comparison group.*

Counterfactual design	*A design in which the understanding of causes is sought by comparing the observed results to those one would expect if the intervention had not been implemented; it implies the use of a control group or comparison group.*
Evidence-based process	*The systematic identification and appraisal of 'best' evidence, based on the level of evidence, its quality, strength, and relevance to a specific topic of interest.*
Generalizability	*The ability to apply the results of a specific study to groups or situations beyond those actually studied.*
Impact	*Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.* [18]
Impact evaluation	*An evaluation that provides information about the impacts produced by an intervention. It can be undertaken of a programme or a policy, or upstream work – such as capacity building, policy advocacy and support for an enabling environment. It goes beyond looking only at goals and objectives to also examine unintended impacts.*
Input	*The financial, human and material resources used in a programme or policy. For example, training materials produced.*
Intervention group	*Also called experimental or treatment group, this is a group of research participants that receives some form of treatment or intervention, i.e. they are exposed to the independent variable. See: comparison group, control group.*
Mean	*A measure of central tendency, commonly called the "average", calculated by dividing the sum of all values by the number of values.*
Outcome	*The intermediate effects of a programme or policy's outputs such as a change in vaccination levels or key behaviours.*
Output	*The immediate effects of programme/policy activities, or the direct products or deliverables of programme/policy activities. For example, the number of vaccines administered.*
(Research) Population	*A group of individuals (or institutions, programs, or other subjects being studied) about whom a researcher seeks to generalize. To generalize about a population, researchers study a sample that is meant to be representative of the population. See: sample.*
Random assignment	*The process of placing research participants/clusters into either intervention or control groups in such a way that each individual/cluster in each group is assigned entirely by chance. That is, each individual/cluster has an equal probability of being placed in each group.*

[18] Organisation for Economic Co-operation and Development – Development Assistance Committee, *Glossary of Key Terms in Evaluation and Results Based Management*, OECD-DAC, Paris, 2010. See http://www.oecd.org/development/peer-reviews/2754804.pdf.

Randomized controlled trials	*A research or evaluation design with two or more randomly selected groups (an experimental group and control group) in which the researcher controls or introduces an intervention (such as a new programme or policy) and measures its impact on the dependent variable at least two times (pre- and post-test measurements). In particular RCTs – which originated in clinical settings and are known as the 'gold standard' of medical and health research – are often used for addressing* ***evaluative research questions****, which seek to assess the effectiveness of programmatic and policy interventions in developmental settings.*
Regression	*A statistical procedure for predicting values of a dependent variable based on the values of one or more independent variables.*
Selection bias	*A bias in the way the experimental and control or comparison groups are selected, resulting in pre-existing differences between the groups that may serve as confounding factors.*
Theory of change	*Explains how activities are understood to produce a series of results that contribute to achieving the final intended impacts. It can be developed for any level of intervention – an event, a project, a programme, a policy, a strategy or an organization.*
Treatment group	*Participants exposed to the independent variable; also called the experimental group.*
Variable	*A clearly defined, quantifiable measure of a quantity that may vary across time for a specific unit of observation, or across space, i.e., between different units of observation.*