

**Impact Assessment on a Shoestring:  
Measuring the Impacts of Forecast-based Financing in  
Resource Limited Settings**

MSc Dissertation by Selby Knudsen

University College London

2021

## **Acknowledgement**

I would first like to thank two people without whom this research would not have been possible, my advisors Dr. Bayes Ahmed and Clemens Gros. Both were always there for me when I had questions and offered all the support I could have asked for. Additionally, Bayes helped me throughout the year as my personal tutor, and he offered tremendous support in all aspects of the MSc program. A special thanks to Clemens, of the Red Cross Red Crescent Climate Centre, who helped me organize my thoughts and ideas and gave me essential insights into the Red Cross Red Crescent Movement.

I would also like to thank those people that took time out of their busy schedules to be interviewed. Without their time, ideas, and insights, I would not have had the data needed to understand the National Societies and formulate the methodologies.

Finally, I would like to thank my friends and family that have supported me throughout this process. Without their encouragement, I could not have produced the results that I did.

## **Table of Contents**

<b>Acknowledgement.....</b>	<b>ii</b>
<b>List of Figures .....</b>	<b>vi</b>
<b>List of Tables.....</b>	<b>vi</b>
<b>Abbreviations.....</b>	<b>v</b>
<b>Abstract .....</b>	<b>1</b>
<b>1. <i>Introduction</i> .....</b>	<b>2</b>
<b>1.1 Research Question and Aims.....</b>	<b>3</b>
<b>2. Literature Review and Background.....</b>	<b>5</b>
<b>2.1 Current Impact Assessment Landscape .....</b>	<b>5</b>
<b>2.1.1 Understanding Impact and Impact Assessment in Humanitarian Work..</b>	<b>5</b>
<b>2.1.2 Issues and Challenges in Impact Assessment.....</b>	<b>6</b>
<b>2.2 Example Impact Assessment Methodologies used in Humanitarian Studies.....</b>	<b>7</b>
<b>2.3 Evidence Specific to Anticipatory Action .....</b>	<b>8</b>
<b>2.3.1 Methods in A-A.....</b>	<b>9</b>
<b>2.3.2 Red Cross A-A Impact Assessments .....</b>	<b>9</b>
<b>2.3.3 Issues with Current A-A Impact Assessments .....</b>	<b>10</b>
<b>3. <i>Methods</i> .....</b>	<b>12</b>
<b>3.1 Evidence Gathering on Methodologies.....</b>	<b>12</b>
<b>3.1.1 Data Gathering .....</b>	<b>12</b>
<b>3.1.2 Synthesis of Methodology Information .....</b>	<b>12</b>
<b>3.2 Interviews.....</b>	<b>12</b>
<b>3.2.1 Study Setting .....</b>	<b>13</b>
<b>3.2.2 Population and Recruitment.....</b>	<b>13</b>
<b>3.2.3 Data Collection .....</b>	<b>13</b>
<b>3.2.4 Data Analysis.....</b>	<b>14</b>
<b>3.2.5 Ethics .....</b>	<b>14</b>
<b>3.3 New Methodology Creation .....</b>	<b>14</b>
<b>4. <i>Results</i> .....</b>	<b>16</b>
<b>4.1 Methodology Descriptions .....</b>	<b>16</b>
<b>4.1.1 General Overview.....</b>	<b>16</b>
<b>4.1.2 Quantitative Studies .....</b>	<b>16</b>
<b>4.1.3 Qualitative Methodologies.....</b>	<b>17</b>

4.1.4	Mixed-Method Methodologies.....	19
<b>4.2</b>	<b>Interview Results .....</b>	<b>21</b>
4.2.1	Current Practices and Challenges.....	21
4.2.2	Capacity at NS .....	22
4.2.3	Important Information to Gather.....	23
4.2.4	Summary .....	24
<b>4.3</b>	<b>Methodology Scoring .....</b>	<b>25</b>
4.3.1	Scoring Details .....	26
4.3.2	Choosing Methodology .....	27
<b>4.4</b>	<b>Selected Methodology.....</b>	<b>29</b>
4.4.1	Elements of Methodology .....	29
4.4.2	SCM For FbF Programs.....	30
4.4.3	Component 1: Developing Impact Model and Model of Success.....	30
4.4.4	Component 2: Quantitative Survey.....	31
4.4.5	Component 3: Qualitative Survey .....	32
4.4.6	Component 4: Analysis and Reporting.....	33
<b>4.5</b>	<b>Modifications .....</b>	<b>33</b>
<b>5.</b>	<b>Discussion.....</b>	<b>36</b>
5.1	Previous Uses of SCM and Modifications .....	36
5.2	Comparison of SCM to Other A-A Impact Assessments.....	37
5.3	Benefits of the Methodology for FbF Interventions .....	38
5.4	Limitations of the Methodology in FbF Settings .....	40
5.5	Limitations of Thesis Study Design .....	40
5.6	Future Directions of Research .....	41
<b>6.</b>	<b>Conclusion.....</b>	<b>43</b>
<b>7.</b>	<b>References.....</b>	<b>44</b>
	<b>Appendix A: Semi-Structured Interview Guides .....</b>	<b>55</b>
	<b>Appendix B: Methodology Scoring Guide .....</b>	<b>60</b>
	<b>Appendix C: Detailed Information on Methodologies Reviewed .....</b>	<b>61</b>
	<b>Appendix D: Additional Interview Quotes.....</b>	<b>82</b>
	<b>Appendix E: Scoring Rationale .....</b>	<b>85</b>
	<b>Appendix F: Success Case Method In-Depth Specifics for FbF Studies.....</b>	<b>90</b>

## List of Figures

Figure 1: Location of FbF programs (Source: German Red Cross, 2020) .....	2
Figure 2: Distribution of methodology types reviewed .....	16
Figure 3: Current practices and challenges for FbF impact assessments .....	21
Figure 4: Comparison of top three methodology scores .....	28
Figure 5: Overview of SCM methodology flow .....	30
Figure 6: SCM modification with non-beneficiary success and failure interviews .....	34
Figure 7: SCM methodology modification with non-beneficiary surveys and interviews .....	34

## List of Tables

Table 1: Quantitative methodologies reviewed .....	17
Table 2: Qualitative methodologies reviewed .....	18
Table 3: Mixed-method methodologies reviewed .....	19
Table 4: Scores for all methodologies .....	25

## Abbreviations

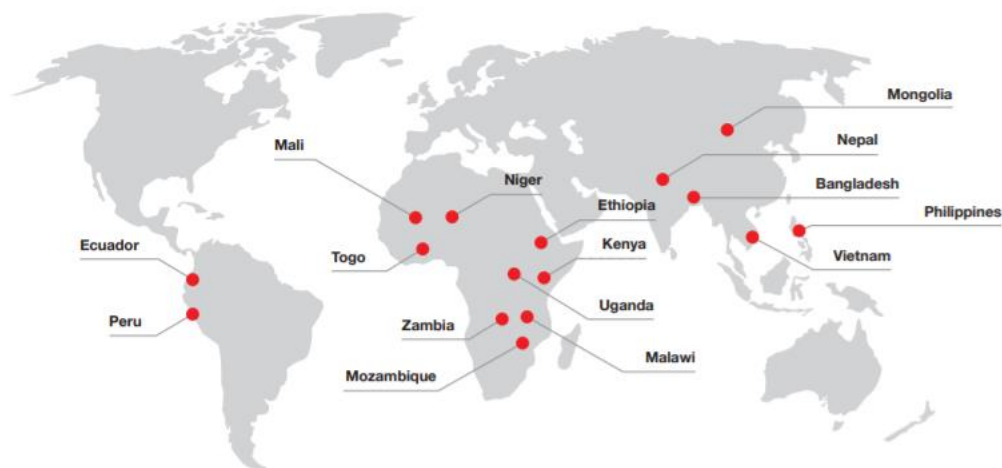
<b>A-A</b>	Anticipatory Action
<b>FbF</b>	Forecast-based Financing
<b>IA</b>	Impact Assessment
<b>NS</b>	National Societies
<b>PIA</b>	Participatory Impact Assessment
<b>QCA</b>	Qualitative Comparison Analysis
<b>QUIP</b>	Qualitative Impact Assessment Protocol
<b>RCRC</b>	Red Cross Red Crescent Movement
<b>RCT</b>	Randomized Control Trial
<b>SCM</b>	Success Case Method
<b>WFP</b>	World Food Program

## **Abstract**

The evidence base of impact assessments in the still relatively novel field of anticipatory humanitarian action is limited, causing a lack of consensus on what is working well and what needs to be changed for these interventions. This scarcity of evidence has been attributed to the capacity and resource requirements of impact assessment methods currently used for anticipatory action, among other reasons. Therefore, the Red Cross Red Crescent Climate Centre, as well as researchers in the anticipatory action sphere, have detailed the need for novel methodologies that can produce robust evidence with lower cost and capacity requirements. This research aims to determine which methodologies work best for impact assessments of the Red Cross Red Crescent Movement's Forecast-based Financing (FbF) anticipatory action interventions. To accomplish these aims, methodologies previously used in the humanitarian and development fields were reviewed. Additionally, FbF stakeholders and employees of various FbF programs, were interviewed to determine current practices and challenges, capacities at different National Societies to conduct impact assessments, and ideas for information to gather. These interviews showed that capacity is variable across different National Societies and that there is a desire to gather qualitative data, which is currently done infrequently during FbF impact assessments. Using the interview results, a methodology was chosen and adapted to address the capacity and resource issues at different National Societies. The methodology chosen is a modification of the Success Case Method, which collects survey data and then uses that data to identify and explore intervention successes and failures. Several additional modifications are also proposed to account for the variability of capacity and allow for flexibility across the different National Societies. The next steps of this research are to discuss the methodology with stakeholders and run several pilot studies to see if it is feasible under real world conditions. If it is successful, this methodology could be used to contribute to the evidence base of anticipatory action going forward.

## 1. Introduction

In recent years, anticipatory action interventions have been gaining popularity as a technique for humanitarian interventions (Weingärtner & Wilkinson, 2019). The Red Cross Red Crescent Movement (RCRC) and the Red Cross Red Crescent Climate Centre (henceforth referred to as the Climate Centre), along with several United Nations organizations, especially World Food Program (WFP), have come up with a novel program meant to develop short-term measures to mitigate negative impact of extreme weather events and decrease the amount of money spent on disaster response (Weingärtner & Wilkinson, 2019). These programs utilize forecasts to organize early actions that can help beneficiaries prepare before a natural hazard strikes an area, therefore hopefully reducing the impact of the disaster (Weingärtner & Wilkinson, 2019; Coughlan de Perez et al., 2015; German Red Cross, 2017). While different organizations have slightly different techniques, the aim is the same: to use forecasts to perform early actions and preparations, which will in turn reduce the impact of the disaster and potentially reduce the overall cost of the response. This report will focus on the Forecast-based Financing (FbF) system that was created by the RCRC with technical leadership from its Climate Centre in conjunction with other stakeholders. Figure 1 shows the countries that have established FbF programs.



**Figure 1:** Location of FbF programs (Source: German Red Cross, 2019)

The RCRC has expanded their FbF programs, and today there are 15 different National Societies (NS) with an established program (German Red Cross, 2020).

There is little published data that shows what impact these programs have on the beneficiaries they are meant to be helping. The impact assessments that have been done by the Climate Centre have been relatively resource and capacity intensive quasi-experimental studies, which have shown mixed results on the effectiveness of the programs (Gros et al., 2019; Gros et al., 2020). These studies require a strong knowledge of statistics, as propensity scores or other complex methods are used to compare beneficiaries with non-beneficiaries (Gros et al., 2019; Gros et al., 2020). The resources and capacity needed for such studies have proved problematic, with more experienced programs supported by more robust NS being able to cope with the logistics required, while more inexperienced NS require a large amount of outside consultation and assistance.

This research aims to address the need for a less resource, capacity, and logistically intensive impact assessment methodology that is still robust enough to aid in the monitoring and evaluation of FbF interventions, so that more data can be gathered on the impact to beneficiary households and the benefits of the intervention. The need for these new methodologies has been an express desire from multiple stakeholders in the anticipatory action sphere (Weingärtner & Wilkinson, 2019).

To accomplish the objectives of finding a more suitable methodology, this report will first outline a comprehensive review of impact assessments in the humanitarian and anticipatory fields and review different methodologies that have been used. It will then present insights from interviews with FbF staff and other stakeholders to better understand the capacity available and impact assessment changes wanted, and use this information to determine what type of methodology would be most appropriate in FbF settings.

## **1.1 Research Question and Aims**

This report aims to answer the following question:

What are suitable methodological approaches for conducting robust FbF impact assessments within given resource, capacity, and time constraints?

### **Research objectives**

- 1) Assess the current evidence landscape on humanitarian and anticipatory action impact assessments and review different methodologies used.



- 2) Record the thoughts of experts on FbF interventions and assess the resources and capacity of different NS to conduct impact assessments.
- 3) Determine a suitable methodology to assess FbF interventions.

### Research aims

This project aims to identify or develop a methodological approach for assessing the effects of forecast-based humanitarian assistance on the disaster-affected population. Specifically, the approach should be as rigorous as possible regarding the assessment of beneficiary impacts while being tailored to the technical capacity and financial means of the average FbF implementing organization.

## **2. Literature Review and Background**

### **2.1 Current Impact Assessment Landscape**

In the development field, countless impact assessments have been done, using different methodologies, to determine if programs are producing their intended outcomes and impacts (Duvendack et al., 2012; Ferrah et al., 2021; Bousquet et. al, 2021; Fuller & Lain, 2017; Duvendack et al., 2014). This wealth of evidence does not exist in the humanitarian field, and even less evidence is available for anticipatory action impact assessments (Proudlock et al., 2008; Weingärtner et al., 2020). Since the 1990s, there has been an increased push for humanitarian organizations to conduct impact assessments on the work that they do (Proudlock et al., 2008). These impact assessments are meant to provide information on what is working and what needs to be changed in humanitarian interventions (Proudlock et al., 2008; DFID, 2012; Leeuw & Vaessen, 2009).

Even with this recent drive to measure the impact of humanitarian action, many final intervention reports do not include an impact assessment. UNICEF has estimated that only 15% of all interventions have an impact assessment as part of its evaluations (Watson, 2008). This chapter will outline what impact assessments are and current practices, the issues and challenges associated with conducting impact assessments for humanitarian work, the evidence base of impact assessments in the anticipatory action field, and the methodological challenges in anticipatory action impact assessments.

#### ***2.1.1 Understanding Impact and Impact Assessment in Humanitarian Work***

Roche (2000, p. 546) has defined impact as 'lasting or significant changes- positive or negative, intended or not- in people's lives brought about by a given action or series of actions.' This definition has been used by multiple different organizations and review papers, so it will be used as the definition of impact for this research (Watson, 2008; Hofmann et al., 2004; Chapman & Mancini, 2008). There are several important aspects of the definition to highlight to understand impact assessments (IAs) in the humanitarian context. First is 'lasting or significant changes.' In humanitarian crises, lasting changes may not be the goal of an intervention. If that is the case, significant changes are assessed to determine impacts. (Hofmann et al., 2004). Secondly, it is important to look at both positive and negative aspects of

interventions (Hofmann et al., 2004; DFID, 2012). The third part of the definition, ‘-in people’s lives brought about by a given action or series of actions,’ reinforces that IAs are specifically meant to look at actions and their effects on people’s lives. In humanitarian interventions, this definition can also be thought of as the preservation of current standards in the face of disruptions (Proudlock et al., 2008).

There are typically two different reasons that IAs are undertaken: learning and accountability (Watson, 2008; Jones et al., 2009; Chapman & Mancini, 2008; Hofmann et al., 2004). Learning studies are conducted to explore what the effects were and why they occurred. These studies are normally done to generate evidence on what does and does not work to improve decision making (Jones et al., 2009). Accountability studies are conducted to measure either upwards accountability to donors or downwards accountability to beneficiaries (Jones et al., 2009; Chapman & Mancini, 2008). This research will focus primarily on learning studies due to their focus on evidence generation.

### ***2.1.2 Issues and Challenges in Impact Assessment***

Due to the nature of humanitarian responses, there are several methodological issues that make humanitarian IAs more challenging than other types of IAs. First, many humanitarian interventions have a lack of baseline data (Proudlock et al., 2008; Watson, 2008; Hofmann, 2004; Bamberger et al., 2004). Humanitarian interventions commonly begin after a disaster, so data collected will not accurately reflect the situation prior to the disaster (Hofmann et al., 2004; Buttenheim, 2010). The lack of baseline data makes IAs challenging, as it makes comparisons to show impact difficult (Proudlock et al., 2008; Watson, 2008). One of the biggest methodological challenges in IA is attributing a change to a specific action. In humanitarian situations, it is possible that many different actions are taking place at the same time. Different organizations are working in the same area, so one specific intervention’s impacts can become obscured by other activities (Watson, 2008; Buttenheim, 2010; Hofmann et al., 2004). There is also issue of impacts not manifesting until later in the timeline, while many donors want an impact report as soon as actions are taken (Bamberger et al., 2004; Proudlock et al., 2008).

To address the issue of a lack of baseline data, several techniques have been established, for example a retrospective baseline or rolling baseline (Proudlock et al.,

2008; Watson, 2008). Retrospective baseline is when participants recall information from before the intervention took place, while rolling baseline is when newly enrolled participants data is used as a baseline to compare the current data of participants that have received the intervention (Watson, 2008). There are also different methods to address the attribution issues, including quasi-experimental studies that have established a counterfactual to determine what would have happened without the intervention (Watson, 2008; Hofmann et al., 2004).

There are many different qualitative and quantitative methods used in humanitarian IAs, but the most common approach is a combination of the two, using mixed methods (Hofmann et al., 2004). When both qualitative and quantitative methods are used, the limitations of one technique are generally reduced when combined with another (Watson, 2008). One analysis of humanitarian IAs conducted since the 1990s showed that 90% involved a mixed methods approach (Proudlock et al., 2008).

## **2.2 Example Impact Assessment Methodologies used in Humanitarian Studies**

While the number of humanitarian IAs is small compared to other development related impact assessments, there are still examples of IAs done (Proudlock et al., 2008; Watson, 2008; Casey, 2015). A meta review of water, hygiene, and sanitation (WASH) interventions in humanitarian crises looked at the quantitative evidence of impact of the interventions (Ramesh et. al, 2015). This review found six studies that used quantitative methods to look at impact, with four using a longitudinal study design without control groups and two using a randomized control trial (RCT) design. The analysis concluded that methodological challenges of the reports analyzed limit the ability to determine association between WASH interventions and health outcomes (Ramesh et al., 2015). A similar meta review was conducted on cash transfers in emergencies. This meta review found five quantitative studies, with three using RCTs and one using a quasi-experimental methodology. This review showed that cash-based approaches can effectively increase food security, although methodological challenges were noted (Doocy & Trappis, 2017).

IAs that use primarily qualitative methods tend to look at community perceptions of the interventions. A meta review of the impact of mental health and psychosocial

support programs following humanitarian emergencies reviewed 15 qualitative impact studies, with all conducting either semi-structured interviews or focus groups. This synthesis shows that qualitative methodologies can provide in-depth information on perceptions of the interventions and changes that have occurred (Dickson & Bangpan, 2015). A review of both quantitative and qualitative impact assessments concluded that quantitative studies could not provide information on why interventions failed, and qualitative information was needed to better understand context (Marston et al., 2013).

Mixed-method approaches vary, but many collect quantitative data that is then triangulated with qualitative interviews or focus groups (Hofmann et al., 2004; Watson, 2008). A review of 113 IAs examining cash transfers in humanitarian situations included 71 mixed-methods methodologies. Many of these included quantitative surveys to collect evidence of impact, and qualitative interviews to better understand the context and challenges of the interventions (Doocy & Tappis, 2017).

### **2.3 Evidence Specific to Anticipatory Action**

While there are a limited number of impact assessments done for humanitarian response programs, there are even fewer done for humanitarian programs that focus on anticipatory actions (A-A) (Weingärtner et al., 2020; Weingärtner & Wilkinson, 2019). The WFP put together a review of the existing evidence of A-A, finding a total of 15 empirical studies on the impact of these programs. One reason for the lack of studies could be the relative newness of the A-A programs, as many pilot programs have not had a triggering event yet (Weingärtner et al., 2020).

Another reason for the lack of evidence could be that impact assessments for A-A encounter the same issues that are seen in humanitarian studies. A-A impact assessments struggle with establishing a counterfactual. There is not always consensus in what the counterfactual should be for A-A, but they are typically either comparing it to a normal humanitarian response or no response at all (Weingärtner et al., 2020; Weingärtner & Wilkinson, 2019). There has been criticism that much of the evidence that is available has focused on what was done, and not on how well it works for whom and what action is needed in a specific context (Weingärtner & Wilkinson, 2019). More evidence is needed in this field to ensure that the programs are having an impact.

### **2.3.1 Methods in A-A**

Of the interventions that have done impact assessments, the methodologies have varied greatly. Anticipatory assessments typically use either quantitative, qualitative, or mixed-methods. Most of the impact assessments in the WFP review were more focused on cost-benefit analysis than the impact on individuals (Weingärtner et al., 2020). Of the ones where the impact on beneficiaries was the main outcome of interest, the most common methodology used was a quasi-experimental study that compared beneficiaries to non-beneficiaries. These studies used statistical techniques, matching, and random sampling. Other designs included collecting qualitative data with no comparison group (Weingärtner et al., 2020). As with humanitarian studies, mixed-methods may be the best approach to IA in the anticipatory action context. A study in Bangladesh that was only able to collect qualitative data emphasized the need for quantitative data to confirm and expand on their findings (Tanner et al., 2019).

A more robust evidence base is needed, both to provide more evidence on the benefits of A-A and to determine whether A-A is more effective than regular humanitarian interventions and how early to provide the intervention (Weingärtner et al., 2020). It has been suggested that developing new methodologies that can deliver robust results in crisis contexts will be important to improve the evidence base for A-A impacts (Weingärtner & Wilkinson, 2019).

### **2.3.2 Red Cross A-A Impact Assessments**

This research focuses specifically on RCRC IAs. Two of the FbF IAs have been published in peer reviewed journals and are well documented. Both employed a quasi-experimental quantitative methodology using propensity scores to match beneficiary and non-beneficiary participants to provide the main body of evidence (Gros et al., 2019; Gros et al., 2020).

The FbF program in Bangladesh targets the most vulnerable populations in the areas that are commonly affected by flooding. In 2017, heavy rainfall caused upstream water levels to rise, triggering the intervention. For this trigger, 1039 beneficiaries were given cash grants (Gros et al., 2019). To assess these cash grants, an impact assessment was done, with the goals of testing whether households that received cash were better able to evacuate, had fewer desperation

sales of assets, consumed better quality food, and had fewer debts than those that didn't receive cash. To do this, a quasi-experimental methodology was used. Four communities that received cash transfers were compared to four communities that did not receive funds (Gros et al., 2019). In addition to the survey, FGDs and KIs were conducted to triangulate the quantitative data (Gros et al., 2019).

In Mongolia, early actions are designed around forecasts for dzuds, which are extreme winter conditions that are preceded by hot, dry summers. In 2017, the risk of dzud conditions were established and the FbF system was triggered. For this trigger, a total of 2000 herders were given a grant of 100 USD each and animal care kits were delivered (Gros et al., 2020). To assess the impact of the FbF early actions, a quasi-experimental study was conducted. This study used multi-stage cluster sampling, where provinces were randomly selected, then districts within the provinces were sampled, and from those districts intervention and comparison households were selected to fill out surveys. Propensity scoring, a statistical technique, was used to match beneficiaries and non-beneficiaries and balance socio-economic differences. Quantitative data on variables such as stock survival, assets, and other financial variables was collected (Gros et al., 2020).

Mongolia showed that there was a significant reduction in mortality of horses amongst those that received the grant (Gros et al., 2020). In Bangladesh, the proportion of households that reinforced their roofs or walls was twice as high in the beneficiary group (Gros et al., 2019). Although some results showed significant differences, additional work is required to fully understand the impact of FbF interventions.

### ***2.3.3 Issues with Current A-A Impact Assessments***

In terms of methodology gaps in the current FbF research, several issues were mentioned in the two papers. In the Mongolia study, researchers noted the need for additional qualitative research to better understand the complexity of the agricultural interventions. The quantitative information was not enough to gain a full picture of the impact (Gros et al., 2020). In the Bangladesh study, researchers noted that methodologies with a longitudinal component would have potentially collected better information on the long-term impacts (Gros et al., 2019). Most importantly, many NS may not have the capacity to conduct such studies, and the funding available may

not cover all costs. One of the reasons there are so few IAs in the humanitarian field is the complexity of quasi-experimental studies (Weingärtner & Wilkinson, 2019). With the lack of capacity and resources required, as well as the methodological issues, several stakeholders in the humanitarian field have expressed a desire for novel methodologies to conduct impact assessments with the resources available.



### **3. Methods**

#### **3.1 Evidence Gathering on Methodologies**

The first step of this report was to do a review of different impact assessment methodologies that are used in the development and humanitarian field. This was done by looking at three different sources: evaluation websites, impact assessment reviews (both in the humanitarian and development field), and individual impact assessments that had been published in journals or by organizations.

Methodologies were excluded if they did not collect data from beneficiaries, if they were only used in business settings and had not been adapted to development or humanitarian fields, if they were not designed to capture impacts, or if there was insufficient information.

##### **3.1.1 Data Gathering**

Once the methodologies were identified, information on each was collected. The information was obtained from four main sources: methodology manuals, methods papers, case studies using the methodology, and evaluation websites. This data collection included gathering information on sampling techniques, comparison groups, data collection needs and techniques, data analysis capacity needed as well as analysis descriptions, type of evidence, and outcome metrics.

##### **3.1.2 Synthesis of Methodology Information**

Once methodologies were chosen and manuals and methods paper found and analyzed, data was synthesized to describe the different methodologies.

Methodologies were first separated by the type of data that was gathered, using the three main categories of data collection: quantitative, qualitative, and mixed methods. Once the methodologies were organized into those categories, information assembled such as sample size, capacity needed to start assessment, data collection information, data analysis information, and evidence gathered was synthesized to create an in-depth description of each methodology.

#### **3.2 Interviews**

In this study, interviews were conducted to determine the capacity of NS to conduct IAs, the resources available, and what information different stakeholders wanted to

get from IAs. Interviews were conducted because very in-depth, detailed information was required to help determine which methodology would be most appropriate, and that information was best gathered from semi-structured, in-depth interviews.

### **3.2.1 Study Setting**

This study was run with input from the Climate Centre to look at different impact assessments that might be feasible for the FbF programs that were being run at multiple NS with input from the Climate Centre. FbF programs are currently active in 15 countries looking at many different hazards. This study focused only on hazards that had a rapid onset, such as floods, and did not investigate the slower onset disasters such as drought. This was because the forecasting and timing of A-A is much more difficult to determine with longer onset disasters. Countries in all stages of developing their FbF programs were considered to determine feasibility, from those that had done impact assessments before to those that did not have any experience.

### **3.2.2 Population and Recruitment**

This study relied on key informant interviews from two different sources, Red Cross employees that worked on FbF at various International Federation of Red Cross and Red Crescent NS, and other key stakeholders in the Climate Centre's network of FbF programs. The aim was to recruit 10 participants with different experience in FbF. These ranged from developing programs and running impact assessments at NS to academics and donors. Key informants were selected purposefully to provide a range of opinions on what is needed in terms of impact assessment for FbF programs. The sample size was chosen because 10 interviews could provide a variety of insights into FbF programs, while being realistic within the time allowed

### **3.2.3 Data Collection**

A total of nine interviews took place over whichever video chat platform was most convenient for the participants. The 10<sup>th</sup> could not be conducted due to scheduling conflicts. All interviews took place over a digital meeting platform. All interviews were conducted in English, and all interviews were recorded with the permission of the interviewees. Semi-structured questionnaire guides and the code book can be found in Appendix A.

The interviewer did not collect any personal information such as age, gender, or position in the organization. Questions differed between those that worked at NS and those that did not. National Society questions focused on questions of previous experience and feasibility of impact assessments, with questions that asked about previous experiences with impact assessments, capacity available at the National Society, resources that might be needed, and what the national society was most interested in learning about their programs. Questions for other stakeholders included gaps in the current evidence, ideas for filling those gaps, outcomes of interest, and general impact assessment ideas.

### **3.2.4 Data Analysis**

Each interview was recorded, either video or audio, and then transcribed. These transcriptions were stored in a secure and encrypted file storage system. Analysis was done in NVivo by the researcher. A thematic analysis was conducted to identify recurrent themes that addressed study questions and then transcripts were coded using those themes (Nowell et al., 2017).

### **3.2.5 Ethics**

The study was approved by the ethics chair of the Institute for Risk and Disaster Reduction at University College London. Participants provided consent, including for video recording. No identifying information was collected.

## **3.3 New Methodology Creation**

The information from the interviews and the methodologies that were reviewed were then combined to try and determine which existing methodology, with modifications specific to FbF programs, best fits the needs, capacity, and resources of the FbF programs at various NS.

To determine which methodology should be used for FbF impact assessments, a scoring system was developed. This system had four components that had been identified as issues to conducting IAs: capacity of national society, complexity of methodology, cost, and stakeholder preferences as determined during interviews. All four components were given a range of scores between 1 and 5, 1 being the most complex, requiring the highest capacity, being the least preferential, or the highest in cost, and 5 being the opposite end of the spectrum. The full description of each

category can be found in Appendix B. After establishing the scoring matrix, each methodology was given a score in all four categories based on the data that had been collected and the average was taken to give them a final score. The top three scoring methodologies were then looked at more in-depth for other factors that may lead to a standout methodology. Once all three were examined, one methodology was chosen based on the characteristics that interviewees determined were most important, and specific FbF related modifications were developed.

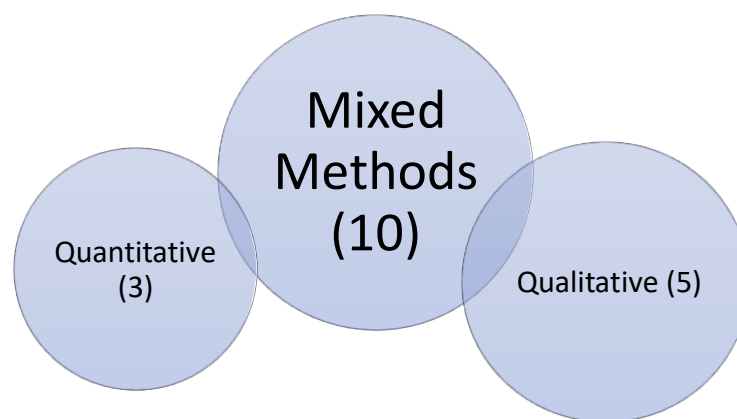
## 4. Results

### 4.1 Methodology Descriptions

Information on the different methodologies was gathered, including sampling techniques, pre-data collection activities, data collection, and evidence gathered. This section will detail the findings.

#### 4.1.1 General Overview

Methodologies fall into three broad categories: quantitative, qualitative, and mixed-methods. Figure 2 shows the breakdown of the categories for the 18 methodologies that were examined.



**Figure 2:** *Distribution of methodology types reviewed*

Most of the methodologies found are mixed-methods, which highlights the trend in IA of gathering both quantitative and qualitative information (Hofmann et al., 2004). More detailed information on all methodologies reviewed can be found in Appendix C.

#### 4.1.2 Quantitative Studies

Quantitative methodologies reviewed are characterized by having large sample sizes, comparison groups, and administering surveys to collect data (Shadish et al., 2001; Gertler et al., 2011; Puri et al., 2014). The methodologies reviewed can be seen in Table 1.

**Table 1:** Quantitative methodologies reviewed

Methodology	Defining features
<b>Randomized Control Trial (RCT)</b>	<ul style="list-style-type: none"><li>• Randomized intervention and control groups, differences in outcome examined (Duflo &amp; Kramer, 2003).</li></ul>
<b>Quasi-experimental Studies</b>	<ul style="list-style-type: none"><li>• Beneficiaries and non-beneficiaries matched and then differences examined (White &amp; Sabarwal, 2014).</li></ul>
<b>Case-Control Study</b>	<ul style="list-style-type: none"><li>• Match those with outcome to those that didn't have outcome (Puri et al., 2014).</li></ul>

Note: References indicate one key reading per methodology

RCTs are considered the gold standard of IAs (Jones et al., 2004). The main tenet of the design is that people are randomized into one group that receives the intervention and one that does not (Duflo & Kremer, 2003; White & Raitzer, 2017). If randomization has been successful, differences in outcomes should be attributable to the intervention (Gertler et al., 2011). Quasi-experimental methodologies use different forms of matching to compare beneficiaries and non-beneficiaries. For example, the Climate Centre studies in Bangladesh and Mongolia use propensity scoring (Gros et al., 2019; Gros et al., 2020; White & Sabarwal, 2014). If done correctly, this methodology can provide evidence on the differences in outcomes that can be attributed to the intervention (White & Sabarwal, 2014).

The case-control methodology is designed to match those that had an outcome of interest to those that did not (Puri et al., 2014). Participants are matched based on demographic and socio-economic variables. This design should be able to make a case for outcomes being attributable to the intervention (Puri et al., 2014).

#### **4.1.3 Qualitative Methodologies**

Qualitative methodologies are characterized by data collection via interviews or focus group discussions. Table 2 shows the defining features of each qualitative methodology.

**Table 2:** Qualitative methodologies reviewed

<b>Methodology</b>	<b>Defining features</b>
<b>Beneficiary Assessment</b>	<ul style="list-style-type: none"><li>• Inquiry into people's beliefs and values about an intervention (Salmen, 1999).</li></ul>
<b>Most Significant Change</b>	<ul style="list-style-type: none"><li>• A collection of stories of change and systematically choosing the most significant (Davies &amp; Dart, 2005).</li></ul>
<b>Qualitative Impact Protocol (QUIP)</b>	<ul style="list-style-type: none"><li>• Work backwards from an outcome to determine how a respondent's life has changed (Copestake, 2014).</li></ul>
<b>General Elimination Method</b>	<ul style="list-style-type: none"><li>• Identifying all possible explanations and then systematically eliminating them (Leeuw &amp; Vaessen, 2009).</li></ul>
<b>Outcome Harvesting</b>	<ul style="list-style-type: none"><li>• Identify outcomes where cause and effect are not well understood, then try to identify causal pathways (Holvand, 2005).</li></ul>

Note: References indicate one key reading per methodology

There are two methodologies that focus more on in-depth case analysis, Beneficiary Assessment and Most Significant Change. The basics of both designs are that individual beneficiaries are interviewed to determine changes that have occurred (Salmen, 1999; Davies & Dart, 2005). Beneficiary Assessment looks explicitly at how beneficiaries believe the intervention has helped these changes occur, while Most Significant Change focuses on what has changed and determining whether the intervention was influential in that change (Serratt, 2009; Salmen, 2002).

QUIP uses both interviews and FGDs to probe for changes in people's lives. The goal is to have the intervention be mentioned as a driver of change without being explicitly probed (Copestake, 2014; Copestake et al., 2018). The General Elimination Method is focused on determining all possible explanations for the outcome and systematically eliminating them until, ideally, the intervention is the only plausible explanation for the outcome. This methodology aims to provide an explanation of causality in the intervention (White & Phillips, 2012; Leeuw & Vaessen, 2009). Outcome Harvesting starts with the outcome and works retrospectively to determine whether the program contributed to these outcomes. This is done by conducting interviews and FGDs (Canto-Blundo et al., 2017; Holvand, 2005).

Qualitative studies help investigators and program staff better understand how the beneficiaries believe the intervention is helping improve their lives. If conducted rigorously, they can provide evidence of attribution (White & Phillips, 2012).

#### **4.1.4 Mixed-Method Methodologies**

Mix-method methodologies have become more widely used in humanitarian IAs in the past few years, as investigators found that having both qualitative and quantitative provides richness in detail. Table 3 shows the mixed-method methodologies that were reviewed.

**Table 3:** Mixed-method methodologies reviewed

<b>Methodology</b>	<b>Defining features</b>
<b>Qualitative Comparison Analysis</b>	<ul style="list-style-type: none"> <li>• Analysis of multiple cases in complex situations to look for patterns that explain why change occurred (Schatz &amp; Welle, 2016).</li> </ul>
<b>Success Case Method</b>	<ul style="list-style-type: none"> <li>• Targeting successes and failures for further interviews to understand why the program worked for some (Brinkerhoff, 2005).</li> </ul>
<b>Case Study</b>	<ul style="list-style-type: none"> <li>• Extensive description and analysis of a situation to gain a comprehensive understanding of what happened (Balback, 1999).</li> </ul>
<b>Contribution Analysis</b>	<ul style="list-style-type: none"> <li>• Gathers evidence to create a contribution story and analyses the theory of change as a contribution story (Mayne, 2006).</li> </ul>
<b>Process Tracing</b>	<ul style="list-style-type: none"> <li>• Analyzes how a potential cause influenced a specific change, using formal tests to the evidence gathered (Anguko, 2019).</li> </ul>
<b>Realist Evaluations</b>	<ul style="list-style-type: none"> <li>• Identifying the underlying mechanisms that explain why outcomes occurred (Westthrop, 2014).</li> </ul>
<b>Collaborative Outcome Reporting</b>	<ul style="list-style-type: none"> <li>• Creating a performance story that presents evidence of how a program has contributed to outcomes (Dart &amp; Roberts, 2014).</li> </ul>
<b>Method of Impact Assessment in Projects and Programs (MAPP)</b>	<ul style="list-style-type: none"> <li>• Group discussions where effects of programs and developments are analyzed using a set program (Neubert, 2010).</li> </ul>
<b>Participatory Impact Assessment</b>	<ul style="list-style-type: none"> <li>• Participatory techniques to determine in changes can be attributed to interventions (Catley et al., 2013).</li> </ul>
<b>Citizen Report Card</b>	<ul style="list-style-type: none"> <li>• Surveys and FGD specifically designed to assess service delivery (Ravindra, 2004).</li> </ul>

Note: References indicate one key reading per methodology



As seen in Table 3, Qualitative Comparison Analysis (QCA), Success Case Method (SCM), and Case Study look in depth at specific cases (Schatz & Welle, 2016; Balbach, 1999; Brinkerhoff, 2005). In QCA, cases are identified that had a particular outcome and compared to cases that did not. The qualitative data gathered from cases is used to create quantitative factor scores, and factors that led to the outcome are determined (Simister & Scholz, 2017). In SCM, the beneficiaries are surveyed to determine quantitative data on outcomes, and then intervention successes and failures are identified and interviewed to examine attributes that led to successes and failures (Piggot-Irvine et al., 2009). Case studies involve gathering as much information about a case as possible, both quantitative and qualitative, to understand what factors contributed to outcomes (Balbach, 1999).

Contribution Analysis, Process Tracing, and Realist Evaluations involve establishing the theory of change, and then gathering data to identify changes and attribute them to the program intervention (Mayne, 2008; Anguko, 2019; Westthorp, 2014).

Contribution Analysis uses previously collected qualitative and quantitative data, as well as conducting interviews, to create a contribution story (Mayne, 2008). Process Tracing determines the theory of change and collects quantitative and qualitative data to document the processes of change and test alternative hypotheses (Anguko, 2019). Realist Evaluations create a context-mechanism-outcome statement to determine what different outcomes occur in different contexts. It looks at differences between subgroups to determine why groups react differently to interventions (Westthorp, 2014).

Collaborative Outcomes Reporting creates a performance story that presents evidence of how the program contributes to outcomes by collecting all available data and gathering new data (Dart & Roberts, 2014; Roughley & Dart, 2009). Citizen Report Card is a methodology that uses FGDs to determine the issues to be assessed. A large survey is then undertaken to collect the thoughts of the community (Ravindra, 2004). Participatory Impact Assessment involves doing interviews and FGDs to determine what changes occurred during the intervention (Catley et al., 2013). Much of the data collected from the FGDs is used quantitatively. MAPP uses group discussions to identify changes and discuss what might have led to the change (Neubert, 2010). Qualitative data is then used quantitatively to rank changes.

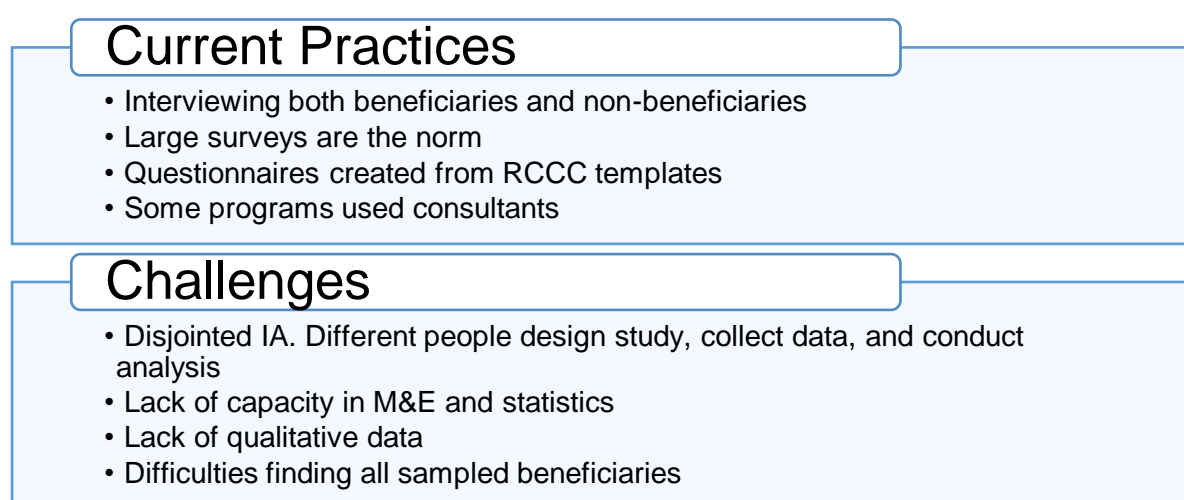
These 18 methodologies were the ones that were most appropriate for conducting impact assessments for FbF programs. Interviews were done to better understand the capacity and resources available at NS, as well as what different stakeholders want in terms of impact assessment.

## **4.2 Interview Results**

A total of nine key informant interviews were completed, with five employees of different NS FbF programs and four other FbF stakeholders. There were four main themes that came up in the interviews regarding FbF programs and IAs. These themes are discussed in depth in this section and more quotes can be found in Appendix D.

### **4.2.1 Current Practices and Challenges**

There were three FbF program members that had been a part of an IA, while two had not. Those that had not were asked about current data collection processes and how they perceive an IA would work with their resources. The interviewees were also asked about current challenges encountered in gathering data. Figure 3 summarizes the current practices and challenges that are faced at the NS.



**Figure 3:** *Current practices and challenges for FbF impact assessments*

As highlighted in the figure, the current practices include doing large surveys with the help of volunteers that interview both beneficiaries and non-beneficiaries. Challenges with current practices include the lack of capacity at NS. Some programs have a robust M&E team with quantitative skills, while some did not.

#### **4.2.2 Capacity at NS**

Capacity at different NS were discussed, both in terms of resources that are currently available, and those that would be needed to conduct IAs. All NS interviewees detailed their prior engagement in the community, and believe it is a large part of their capacity to do IAs. One interviewee believed that these ties were the reason that participants willingly shared information, saying:

*“all of the district level units have some sort of program, an ongoing program.... And that is really helpful to make people aware of the host national society. That is why they share the info.”*

Apart from community ties, the vast volunteer network is a major capacity that can be used to conduct IAs. All NS interviewees said that any surveys conducted have used volunteers to collect data. Most have said that it is very easy to get volunteers; previous assessments have used between 20 and 100 volunteers.

The capacity of NS in terms of human resources varies drastically, an observation that was detailed by the NS staff and several of the stakeholder. One stakeholder said that IA was difficult because “many of these NS don’t have teams of M&E (Monitoring & Evaluation), but maybe a person that takes care of PMEL (Planning, Evaluation, Monitoring, and Learning), so that is something we need to change.” While this is the case in some NS FbF programs, others have a more established team. One interviewee detailed how their M&E team would help with IA:

*“Next time we will be engaging our PMEL department on any impact assessment done...And last month we conducted a simulation on heat wave, and we engaged the M&E department for monitoring and evaluation purposes as well as post-distribution monitoring.”*

One interviewee detailed their lack of capacity in statistical analyses, saying “if we had to do quantitative analysis or cost-benefit analysis or quantitative surveys we would need those skills.” Other NS have a higher capacity, where the statisticians can do “data collection frameworks, sampling design, field testing, as well as data analysis.”

### **4.2.3 Important Information to Gather**

Another theme of the interviews was information both stakeholders and NS FbF program staff wanted to collect while doing these assessments.

#### **Qualitative Data**

Almost all interviewees said that qualitative data was essential to collect, because as one stakeholder put it, “we know that not everything can be measured with statistics, there might be a weird reason that you have that finding that might have to do with the causality.” Another stakeholder had similar views, saying “I think in all cases you’re going to want to have some qualitative interviews.”

NS FbF staff had the same opinion as the stakeholders, with one saying:

*“Without qualitative data, they cannot express their expression. So we need to keep a few qualitative information. How they feel these days, what feels different. This cannot be expressed in quantitative [data].”*

#### **Outcomes of Interest**

Interviewees also detailed important outcomes of interest to collect during any impact assessment. One stakeholder was interested in “evidence on how to pick the action that will have the most benefit to the people that will need it the most.”

For the FbF staff, the specific outcomes of interest varied between the different interviewees. One interviewee said: “I would like to see how our support benefits them to save their livelihoods and lives.” Two other interviewees agreed. Others had more general ideas of what outcomes were needed. One interviewee summed it up well, saying:

*“What we would like to demonstrate is that if we do an anticipatory action we are making a difference and we are clearly reducing the impact that we are trying to mitigate and that we are making the overall response and recovery less needy.”*

All interviewees wanted to see if their FbF programs were able to make a difference in the lives of beneficiaries.

#### **How the program is working**

Stakeholders also wanted to collect information on how the program was working. One stakeholder wanted future IAs to look at program design to determine “why and what could be done differently to make sure they [beneficiaries] weren’t worse off.” Other interviewees were more interested in “if we achieved what we wanted to achieve.”

One interviewee detailing exactly what was needed in the IA to show that the program was working, saying:

*“Now we need to provide proof that indeed if we are providing this [house stabilization] kit in advance we will make a clear difference at the local level and we will be able to minimize the impacts of the typhoon winds.”*

Others want to use these impact assessments to determine whether the early actions in their FbF programs are the most appropriate, saying that “an impact assessment would be appropriate. Because the early action seems logical and correct, but maybe some people might ask if it is the right design.”

### **Other data**

There were many mentions of other types of data interviewees wanted collected during IAs. Some wanted impact assessments to include “different ways to indicate impact from the indicators we have,” and “efficiency of the fund.” One stakeholder was more interested in a specific aspect of the program, saying:

*“If you could answer the question, when is cash not going to be a good idea, then you could get people started on using cash in other situations.”*

Some interviewees believe that “the main point of FbF is not whether we are addressing the impact, but whether it is cost beneficial and viable in terms of value for money to do these actions before,” and wanted assessments to focus on that as well.

### **4.2.4 Summary**

The most important information in determining which methodologies to use are capacity and preferences for information to be gathered. All interviewees agreed that collecting qualitative information is very important in any IA. Some detailed difficulty finding some participants in the sampling, but liked having comparison groups. In

terms of capacity, all programs rely on a vast network of volunteers to administer surveys, and that could be tapped into to conduct IAs. Some programs have a strong M&E and statistical capacity, while others lack that.

### 4.3 Methodology Scoring

Each methodology was assessed for the four categories and given a score between 1 and 5 for each category, 1 being the worst (most capacity, least preferable, high complexity, high cost) and 5 being the opposite. Table 4 shows the results of the scoring. Details on the rationale behind each score can be found in Appendix E.

**Table 4:** Scores for all methodologies

	<b>Capacity Required</b>	<b>Preference</b>	<b>Complexity</b>	<b>Cost</b>	<b>Total Score</b>
<b><i>Quantitative</i></b>					
<b>RCT</b>	1	2	1	1	1.25
<b>Quasi-Experimental</b>	1	2	2	2	1.75
<b>Case Control Study</b>	2	2	2	2	2
<b><i>Mixed-Method</i></b>					
<b>Contribution Analysis</b>	3	3	3	4	3.25
<b>Process Tracing</b>	3	3	3	4	3.25
<b>Realist Evaluation</b>	2	4	3	4	3.25
<b>Qualitative Comparison Analysis</b>	3	4	3	5	3.75
<b>Case Study</b>	3	2	3	4	3
<b>Success Case Method</b>	3	4	4	4	3.75
<b>Collaborative Outcome Reporting</b>	2	3	2	3	2.5
<b>Citizen Report Card</b>	3	2	4	2	2.25
<b>Participatory Impact Assessment</b>	4	4	3	3	3.5
<b>MAPP</b>	3	3	4	2	3
<b><i>Qualitative</i></b>					
<b>Beneficiary Assessment</b>	3	2	4	4	3.25
<b>Most Significant Change</b>	4	2	4	2	3
<b>QUIP</b>	4	2	4	2	3
<b>General Elimination Method</b>	3	2	3	4	3
<b>Outcome Harvesting</b>	3	2	4	4	3.25

### **4.3.1 Scoring Details**

#### **Capacity**

The capacity score was based on capacity that programs would need to conduct an assessment using the methodology. Those that scored lowest on capacity, RCT and quasi-experimental, both require very complicated statistical analysis, as well as sampling and matching or randomization (White & Sabarwal, 2014). Some programs may have this capacity, but many would find this prohibitive to conducting IAs. The two that scored the second lowest, Realist Evaluations and Collaborative Outcome Reporting both require extensive work before data collection can start, including designing and hosting workshops, developing analytical frameworks, collecting large amounts of data, in addition to questionnaire design and data analysis (White & Phillips, 2012; Dart & Roberts, 2014). Those methodologies that scored higher have less capacity required, with less statistics or less activity before data collection.

#### **Preference**

The preference scores consider information gathered from the interviews about what the NS want from IAs. This includes aspects like both quantitative and qualitative data, comparisons between beneficiaries and non-beneficiaries, and difficulties of finding all the sampled participants in the more complicated sampling designs. Most of the quantitative and qualitative designs did not receive very high scores, because neither included both data types. Those that received the highest scores had more than one aspect of their methodology that interviewees preferred, most often including both quantitative and qualitative data as well as looking at a comparison group.

#### **Complexity**

Complexity scores come from the number of steps involved in the methodology, complicated data gathering techniques, complicated data analysis, or other complications. Quantitative studies have high complexity scores because require large sample sizes, matching or randomization techniques, and complex data analysis (Duflo & Kremer, 2003). Other methodologies that have high complexity are Collaborative Outcomes Reporting, which has six steps, all involving complications (Dart & Roberts, 2014). Many of the more qualitative methodologies have lower

complexity scores, as sampling and data analysis is generally less complex. Success Case Method has a low complexity because research has suggested that it is a straightforward approach, with a short survey on a small sample to determine successes and failures and then a few qualitative interviews per group (Bell & McDonald, 2006).

### **Cost**

Cost scores were the most difficult to determine, as most papers did not include information on cost. It has been noted, both in interviews and research, that RCTs and quasi-experimental studies are expensive (Hulme, 2000). This can be for several reasons, including the large number of volunteers that are needed to collect data and the potential for having to hire a consultant. Other methodologies, such as QUIP, specifically mention that questionnaire design and data collection need to be done by consultants or outside investigators. Hiring someone to do that will make the methodology more costly (Copestake, 2014). Any methodology that requires a large sample size was also given a high-cost score, as having many volunteers in the field increases cost.

#### ***4.3.2 Choosing Methodology***

After scoring all methodologies on the four factors, the average was taken to determine the total score. Qualitative Comparison Analysis (QCA), Success Case Method (SCM), and Participatory Impact Assessment (PIA) had the highest overall scores. Figure 4 gives an overview of the different categories these methodologies were assessed on, and the rationale behind the scores they received.



Participatory Impact Assessment	Success Case Method	Qualitative Comparison Analysis
<ul style="list-style-type: none"> <li>•<b>Capacity (4):</b> Qualitative data gathering and analysis skills needed, most questionnaires already developed</li> <li>•<b>Preference (4):</b> Collects qualitative and quantitative data, but no large surveys. Gets info on both beneficiary and non-beneficiary</li> <li>•<b>Complexity (3):</b> Logistically complicated, could be large sample sizes for interviews</li> <li>•<b>Cost (3):</b> Depending on sample size it could be large</li> </ul>	<ul style="list-style-type: none"> <li>•<b>Capacity (3):</b> Some sampling and questionnaire development required, basic quantitative and qualitative analysis</li> <li>•<b>Preference (4):</b> Collects survey data and qualitative data, makes a comparison (success vs failure), does not collect info on non-beneficiaries</li> <li>•<b>Complexity (4):</b> Fairly simple, a survey and then follow-up interviews on a very limited sample</li> <li>•<b>Cost (4):</b> Moderate/low cost. Need to pay volunteers for survey, but otherwise low cost</li> </ul>	<ul style="list-style-type: none"> <li>•<b>Capacity (3):</b> Knowledge to develop theory of change, M&amp;E capacity needed to identify cases and factors, develop questionnaires, qualitative data skills</li> <li>•<b>Preference (4):</b> No survey data, but does collect qualitative info and has a comparison group (not beneficiary vs non-beneficiary)</li> <li>•<b>Complexity (3):</b> Start up phase is complex, data gathering less complex, potential to re-visit</li> <li>•<b>Cost (5):</b> Lower cost because the sample size is so small</li> </ul>

**Figure 4:** Comparison of top three methodology scores

All methodologies had similar capacity scores, with SCM and QCA requiring a slightly higher capacity than PIA. This is because PIA has the questionnaire frameworks developed, so questionnaire development is easier, and they can focus on sampling and qualitative data collection and analysis, where the others need to develop surveys (Catley et al., 2013). For Preference, all three methodologies had the same score of 4. This is because all three managed to fulfill most of what was required, but each was missing one element. SCM does not collect beneficiary vs. non-beneficiary comparison data, whereas PIA does not collect survey data (Catley et al., 2013; White & Phillips, 2012). For complexity scores, PIA and QCA received similar scores, while SCM received a higher score. Both QCA and PIA received the lower score because they are logistically more complicated or have complex dynamics in the startup phase (Catley et al., 2013; Schatz & Welle, 2016). SCM was given a 4 because it involves short surveys with small sample sizes, which all programs do and generally find easy when complicated sampling is not involved, and then a small number of interviews. All research done using this technique has commented on how quick and easy it was, which factored into the score (Brinkerhoff, 2005; Bell & McDonald, 2006; Medina et al., 2015; Coryn et al., 2009). PIA was

given the lowest cost score, because of the potential for a very large number of interviews. This would require a lot of trained volunteers, which could be costly. SCM was given the second highest cost score because volunteers would be needed to collect survey data. QCA will have the lowest cost because the sample size is much lower than the other methodologies.

#### **4.4 Selected Methodology**

SCM was selected as the methodology for FbF IAs. This was chosen because it is a relatively quick and easy methodology, that still provides rigor and meets many of the preferences that were indicated by the interviews. It was chosen over QCA, which had the same overall score, because SCM collects survey data. There are some slight modifications made to fit the specifics of the FbF programs, as well as further modifications that can be made to this methodology that would meet almost all preferences.

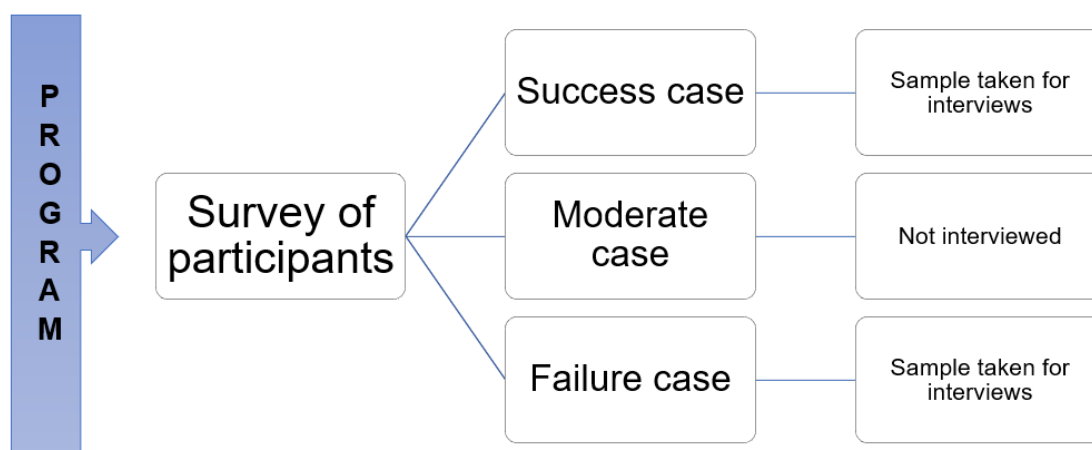
This section will discuss the elements of the methodology, how the methodology will work in the context of FbF, and detail modifications for higher capacity NS.

##### **4.4.1 Elements of Methodology**

The SCM methodology has four basic components: developing an impact model and model of success, using the method to develop a survey to identify successes and failures, conducting in-depth studies of successes and failures, and reporting and analyzing all the findings (Piggot-Irvine, 2009; Clinton et al., 2007). The impact model explains how the intervention is supposed to produce its impacts. Success models are determined to figure out what will be classified as a success (Coryn et al., 2009; Bell & McDonald, 2006). Once these two aspects are developed, a survey is administered to either all or a sample of participants. The survey should be designed to determine the classification of cases as success, moderate, or failure and collect basic intervention information (Olson et al., 2011).

Once all information is gathered, cases are categorized into successes, failures, and moderates based on the success model developed. A sample of the successes and failures are then randomly sampled and selected for in-depth interviews to determine why the success cases succeeded and why the failures failed (Piggot-Irvine et al., 2009). The point of these interviews is to probe for possible explanations of success,

with one possible reason being the intervention. Finally, the findings and conclusions are written up and communicated. This has been regarded by users as an easy, quick, and inexpensive way of conducting impact assessments. Figure 5 gives a visual representation of the steps involved in the SCM methodology.



**Figure 5:** Overview of SCM methodology flow

#### **4.4.2 SCM For FbF Programs**

This section details how this methodology will work within the context of FbF programs, then suggest modifications that programs can make. A more in-depth look at all aspects of the SCM methodology for FbF programs can be found in Appendix F.

#### **4.4.3 Component 1: Developing Impact Model and Model of Success**

For FbF programs, the focus of the assessment is typically to determine the effectiveness of forecast-based early actions to help vulnerable communities take early actions and assess the early actions' impact in preventing or reducing the negative impacts of hazards to health, well-being, assets, and livelihood (Gros et al., 2019). Therefore, the focus of the SCM should be to assess whether the FbF early action programs achieved their goals. The basic impact statement for FbF is the following: By providing early actions before a forecasted hazard, beneficiaries should be able to prevent or reduce negative impacts such as loss of life and livelihoods.

The model of success will vary between interventions because it depends on the early actions chosen, the hazard, and the impacts they are meant to address. To give an example, the intervention in Bangladesh distributes cash grants before

flooding to help beneficiaries reduce the negative impacts of flood on health, livelihoods, and assets (Gros et al., 2019). For this case, the model of success would be a beneficiary who used the grant money to take preparatory action that reduces the impact from the flooding on health, livelihoods, and assets.

#### **4.4.4 Component 2: Quantitative Survey**

The quantitative survey will gather basic information on the intervention and impact to find success and failure cases. Using the Bangladesh example, the survey will need to ask questions about conditions prior to receiving the grant, what was done with the grant, and conditions after the flood in relation to health, livelihoods, and assets.

##### *Sampling and Sample Size*

Many of the studies that have previously used this methodology have administered the survey to all participants, because they had a small number of participants (Coryn et al., 2009; Clinton et al., 2007). The FbF programs have much larger number of participants, for example 1059 people were given the cash grants in Bangladesh (Gros et al., 2019). Therefore, a sample of participants needs to be taken. Sampling can be done in several ways, depending on the geographical area where the trigger occurs. If the area is small, one stage random sampling can occur. If the geographical area is larger, it may be better to do two-stage cluster sampling, where districts are first randomly sampled, and then individual beneficiaries are randomly selected from the sampled districts (USAID, 2021; SMART, 2012; OXFAM, 2019). In studies used to assess WFP interventions on nutrition, they found that if there is limited data on the variables that are being assessed, a sample size of 150 to 250 should be sufficient (WFP, 2009). According to Israel (1992), with a population size of 1000, a sample size of 91 would result in an error of  $\pm 10\%$  and a sample size of 286 would result in an error of  $\pm 5\%$ . Therefore, the upper bounds of the WFP sample size would have an error of approximately  $\pm 5\%$ . If a NS has the capacity to conduct sample size calculations, that could be done instead.

##### *Administration of Surveys*

Interviewees all mentioned that volunteers had a lot of experience in administering surveys, so they should be used. Administration of this survey will require a day long

training for volunteers. In past IAs, there have been issues of not being able to find beneficiaries, so backups will need to be identified in advance.

### *Analysis*

After the surveys are administered, the data will need to be analyzed to find the success and failure cases. This will require someone with basic knowledge of statistics to look at the data. In prior studies, there have been specific cut offs for numbers that have been considered a success and failure, and all other cases are classified as moderate (Coryn et al., 2009). The same should be done in this case. This survey data should also be used to provide quantitative data on the intervention.

#### **4.4.5 Component 3: Qualitative Survey**

The next step of this methodology is to sample a few of the successes and failures for interviews to determine what factors led to the success and what led to the failures. Different studies have done this differently, with some choosing to only look at successes, but for FbF it is important to have a comparison and to see why things did not go well (Brinkerhoff, 2005).

### *Sampling and Sample Size*

The number of successes and failures has varied, but there are usually between 2 and 6 people from both groups interviewed (Coryn et al., 2009; Clinton et al., 2007). As FbF studies have larger beneficiary numbers than previous studies done, the sample size should be a bit larger to help understand different reasons for success and failure, between 6 and 10 in each group.

### *Questionnaire Development and Administration*

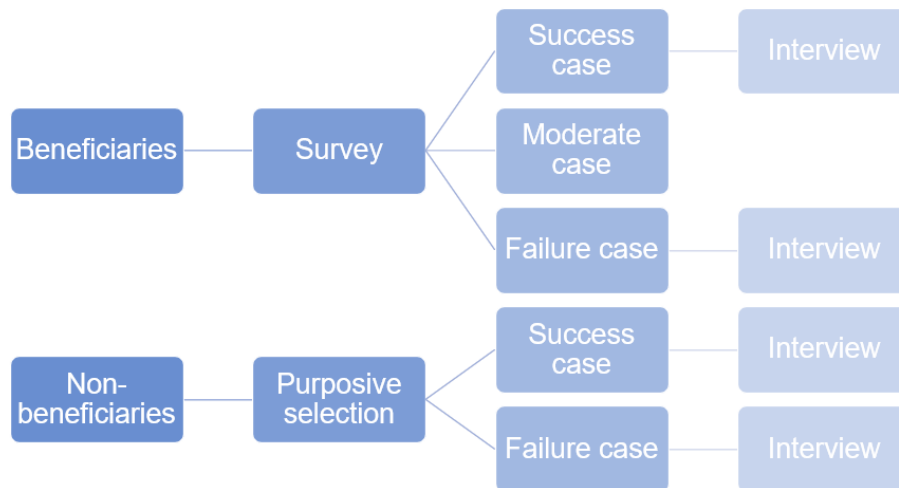
Qualitative questionnaires will be developed differently for successes and failures. The goal of these questions will be to allow an open-ended conversation about why successes were successful and why failures failed. In the example of Bangladesh, this would include questions on how the grant money was spent, what challenges occurred when preparing for floods, and what occurred to the beneficiary after the flooding (Gros et al., 2019). Questionnaires will be administered as semi-structured interviews. Staff members or volunteers that are trained in qualitative interviews can conduct the interviews.

#### **4.4.6 Component 4: Analysis and Reporting**

Qualitative analysis will focus on themes from interviews that illustrate why success cases became successes and why failures failed. This will help determine if there is a pattern to success and failure, and what some of those conditions are (Piggot-Irvine et al., 2009). The results are then written up into a report. Most of these reports are typically written as “success stories,” but in the case of FbF, the descriptive statistics and the failure stories should also be included to provide context and detailed analysis of failures (Brinkerhoff, 2005). Looking at failures will hopefully provide evidence of changes that can be made to improve the interventions.

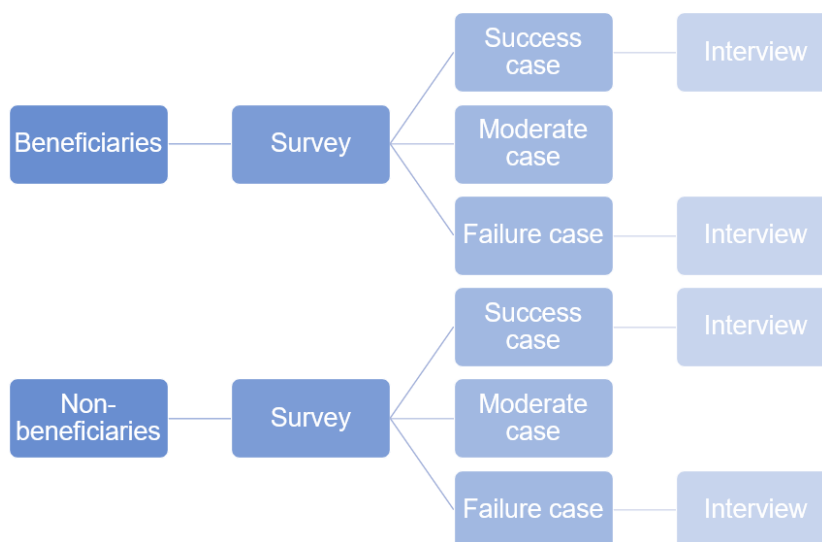
#### **4.5 Modifications**

Depending on the capacity and funding available at different NS, there are several modifications that can be made. First, many interviewees mentioned the desire to also look at non-beneficiaries. If the capacity and funding is available, non-beneficiaries could be surveyed and several success and failure cases should be interviewed. This will give a brief picture of the characteristics of the non-beneficiaries and some of their success and failure factors. The diagrams below show the two different options for non-beneficiary groups. First, if limited capacity is an issue, non-beneficiaries can be purposefully sampled from people that did and did not do well following the disaster to conduct success and failure interviews (Figure 6). This would reduce the capacity needed to conduct surveys but would still provide some non-beneficiary information.



**Figure 6:** SCM modification with non-beneficiary success and failure interviews

The other option, for the NS that have greater capacity, would be to survey both beneficiaries and non-beneficiaries (Figure 7). The sampling could follow the WFP (2009) guidelines, which needed 150-250 participants per group. Alternatively, if capacity and funding is available, NS could conduct their own sample size calculations. After the survey, success and failure cases for both beneficiaries and non-beneficiaries should be interviewed. Fewer participants could be surveyed, but that would limit the representativeness and increase sampling error. A minimum of 100 participants per group has been suggested (Singh & Masuku, 2014).



**Figure 7:** SCM methodology modification with non-beneficiary surveys and interviews

One study that used a modified version of SCM decided to include a time series element to the design, re-surveying participants to see if participants moved categories at three different time points (Coryn et al., 2009). This could be a modification to FbF impact assessments if a program had the capacity and wanted more long-term data on impacts.



## **5. Discussion**

### **5.1 Previous Uses of SCM and Modifications**

Success Case Method was first developed by Brinkerhoff in 2005 as a methodology to assess the impact of trainings for businesses and other entities. Brinkerhoff (2005) stipulated that the traditional quantitative methods under-represent that best case and over-represent the worst when dealing with means. This technique tries to highlight the best and worst cases in the intervention to see what is working and what is not (Olson et al., 2011). The SCM provides an in-depth look at household or individual stories as well as general information on most participants. It evaluates both short- and medium-term outcomes and their relationship to outputs, and can be used to identify possible mechanisms for positive and negative impacts (Coryn et al., 2009; Bell & McDonald, 2006; Piggot-Irvine et al., 2009). It is also a relatively quick way of doing an impact assessment, as typical SCM studies use very limited surveys to determine success cases. While SCM is traditionally used to assess organizational interventions, the methodology has been adapted for several different circumstances.

Bailey et al. (2006) used SCM to evaluate the impacts of a healthy housing program in New Zealand. In this case, the households were chosen by providers based on available database information. In total 9 from each suburb were chosen, with 3 successes, 3 failures, and 3 complex cases. The complex cases were cases that had aspects of both success and failure (Bailey et al., 2006; Clinton et al., 2007). Another SCM modification occurred for the evaluation of an NGO program. Instead of only one sampling point, they found the success and failure cases at three different time points during the study period (Coryn et al., 2009). This study started out with the more traditional SCM method, where all 75 people that had completed the program to reduce homelessness were surveyed and placed into high, middle, and low groups based on the program definition of success. A random sample of 1/3<sup>rd</sup> of the participants in both high and low categories were selected for interviews. This whole procedure, including the survey, was done at two other time points to see if participants changed classes over time (Coryn et al., 2009). This design would satisfy the criticism in the Bangladesh paper that wanted more follow-up data.

Another study used the SCM methodology to assess psychosocial capacity building following earthquake disasters in China. A total of 14 women were given a survey, and the four that were interviewed as 'success cases' because they scored highest on the survey (Sim et al., 2019). The survey was used to collect demographic information on the participants as well as determine successes. FGDs were conducted before the survey to gather additional information on psychosocial capacity building. Additional SCM modified studies include two studies on using medical education to enhance research and practice (Olson et al., 2011; Medina et al., 2015). In Olson et al., only success cases were interviewed, and Medina et al. interviewed all participants, not just successes and failures. These five studies show that SCM can be successfully modified. There are limited publications on SCM methods in general, and only one found using SCM in humanitarian situations, but with modification it could be a suitable FbF impact assessment methodology (Sim et al., 2019).

## **5.2 Comparison of SCM to Other A-A Impact Assessments**

One of the most important aspects of IA is to determine what is working and what is not working, so that changes can be made to address things that are not working. This methodology provides an excellent opportunity to do so. By looking in-depth at the failure cases, patterns should be found that can illustrate what has not worked for people. Similarly, the success cases can help determine what is working well, for whom, and why (Brinkerhoff, 2005; Ball & McDonald, 2006).

As there are only a limited number of impact assessments done for anticipatory action, there are limited methodologies to compare this proposed SCM model to. If using the modifications, it seems to solve several of the issues that came up in the RCCC publications on FbF impact assessments in Mongolia and Bangladesh, namely the lack of qualitative data and the potential to look at longer term impacts (Gros et al., 2019; Gros et al., 2020). Additionally, according to multiple stakeholders interviewed, these quasi-experimental studies were too resource intensive and required capacity that the programs did not have available. The relative simplicity of this design compared to quasi-experimental designs may help fix that.

Other impact assessments that have been done on anticipatory action interventions have used a variety of methodologies. FAO has used a mix of historical data,

surveys, and interviews to determine the impact of their interventions (FAO, 2018; FAO, 2018a; FAO, 2019; FAO, 2020). While their reports did not provide extensive detail on the methodology, the results showed that they used in-depth case studies to understand the interventions impact on individuals, without the use of a control group (FAO, 2019; FAO, 2018). This shows that IAs that do not use a control group still provide valuable information on impacts on beneficiaries. Therefore, if capacity is only available to do the original SCM design, valuable information on the intervention and beneficiaries can still be collected.

In Uganda, an IA was done on flood anticipatory action, which used comparison groups and representative samples to assess impact. This helped the team learn which actions produced the expected impacts, because they could compare it to communities that did not receive the actions (Jjemba et al., 2018). An IA of anticipatory action in Bangladesh used a more theory-based approach, determining impacts based on examining a set of assumptions within the project's logic chain, similar to some of the theory-based methodologies described in the methods section (Tanner et al., 2019). They interviewed beneficiaries and a comparison group as part of that process. Using the modified version of the SCM involving beneficiary and non-beneficiary comparisons can provide similar data.

### **5.3 Benefits of the Methodology for FbF Interventions**

The biggest issues with the previous methodologies used for FbF interventions were the fact that they were beyond the capacity of many of the NS to conduct on their own, and that they were not collecting qualitative information to better understand the nuances of the intervention impacts (Gros et al., 2020). Interviewees expressed concern that IAs were being done with little input from NS due to limited capacity. The biggest benefit of this methodology is that even NS with more limited capacity can be drivers of their next IA. With the modifications, the proposed methodology can be tailored suit each NS needs.

For FbF programs, there are two major modifications that have been suggested. While the SCM without modifications is adequate for NS with limited resources, these modifications can be used to gather more data if resources allow. The first is to add a non-beneficiary group to both the survey and failure/success aspects of the methodology. This would involve sampling non-beneficiary communities, conducting

surveys, and then analyzing the data to determine successes and failures. These successes and failures then need to be interviewed to determine reasons for successes and failures. The sample size can be determined by the resources available, but should be representative enough to make some inferences about the beneficiary and non-beneficiary community. This modification satisfies the necessity expressed by several interviewees to compare beneficiaries and non-beneficiaries.

The procedures for non-beneficiaries will be the same as what is proposed for beneficiaries. If information is collected on both beneficiaries and non-beneficiaries, some comparisons can be made. An interviewee gave an example of an FbF intervention in Uganda where beneficiaries and non-beneficiaries were interviewed and from that the study team learned that although they gave beneficiaries shovels to build latrines, the non-beneficiaries already had shovels because it was a farming community. The modification to the SCM model that has been proposed would pick up on some of the same insights.

The other modification is to follow the design of the Coryn et al. (2009) study, where the entire survey and success/failure interviews are done at several time periods. This modification would be beneficial for programs that are interested in determining long term impacts, as well as seeing how things change over time.

The success case interviews can also provide very valuable insight into what is working well with the program, and if it works best for a certain type of beneficiary (Piggott-Irvine et al., 2009). It will reveal factors that lead to success, which NS can use to tailor the FbF programs to help more people. This level of information cannot be gathered from the quantitative surveys alone.

Besides collecting qualitative information to measure the impact of these interventions, this different methodology has the added benefit of requiring less capacity than the quasi-experimental studies that have been conducted previously for FbF interventions (Bell & McDonald, 2006). Because there is no comparison between beneficiaries and non-beneficiaries in the basic design, the sample size can be smaller, and the sampling techniques less intensive. The statistical capacity required for analysis is also much less intensive for SCM methodology than for the quasi-experimental methodologies that have been used previously, because no comparisons or matching need to take place in the unmodified version. It does

require the capacity for some qualitative analysis, but that is less specialized than statistical analysis and can be taught to anyone on the staff.

#### **5.4 Limitations of the Methodology in FbF Settings**

While this methodology addresses several of the issues that have occurred because of the quasi-experimental study, there are still limitations to this design. The first is that although this is a simplification of the sampling design from quasi-experimental studies, it does require some sampling skills. This should be fine for most NS, but those without any statistical capacity may still struggle.

The biggest limitation to this design without modifications is that it does not collect information on non-beneficiaries. This is something that most interviewees said they wanted out of a methodology. While this is an issue that needs to be addressed, this is only an issue for those NS with limited capacity and resources. For those that do have a larger capacity, they can use one of the modification techniques discussed in the results section to gather information on non-beneficiaries.

#### **5.5 Limitations of Thesis Study Design**

While there are real benefits to this research, there are also several limitations regarding timing and availability of data. First, it may have been better to interview a representative from all the FbF programs that have been developed. This would have given an even better understanding of the range of capacities available and experiences at the NS. This was not done due to a lack of time available for the research. Although this is a limitation, the nine interviews that were done gave a good understanding of what is being done and what is needed going forward. Additionally, the interviews with National Society employees gave a very good range of capacity, from a Society with very strong statistics and M&E team to a society that has very limited statistical capacity. Another limitation of the study was that only those NS members who had designed the FbF protocol were spoken to for interviews. It would have been a good idea to discuss these methodologies, capacity, and resources with members of the M&E team to get their perspectives.

Finding methodologies that had previously been used in humanitarian disaster response settings was difficult. As mentioned previously, development studies are much more common than humanitarian impact assessments, therefore most of the

methodologies were used in development studies. Additionally, experimental or quasi-experimental studies are still considered to be the gold standard. As these are difficult to conduct in humanitarian situations, for reasons discussed in the literature review, there is limited research in general on humanitarian and anticipatory action interventions. Because of the lack of studies, most of the methodologies chosen were used in development studies and there were few examples of their use in humanitarian research. Modifications will have to be made to most methodologies found to be appropriate for FbF interventions.

One other limitation of the study was that if a methodology did not have a case study with a really well detailed methods section, a methodology guide, or methods paper, it was difficult to get the information required to adequately score the methodology. There were a few impact assessments that were done for various interventions that did not have a detailed enough methods section to include in this research, although they might have been good candidates if more information had been available. This lack of methodological information has been documented in many humanitarian IA meta-analyses (Ramesh et al., 2015; Doocy & Trappis, 2016)

## **5.6 Future Directions of Research**

There are several different directions to take this research going forward. The next step would be to get buy-in on the new methodology from several NS. They should be given all the information about the methodology and the modifications, and then interviews should be conducted to determine if they think this new methodology would be preferable to quasi-experimental studies. Box 1 shows the five studies that have been used to demonstrate different SCM modifications as well as the original methodology design.

There have been five modified SCM studies identified that were used to assess a variety of programs, from health education to psychosocial capacity building. These studies can provide good insight for anyone considering adopting the methodology. Brinkerhoff (2005) introduces the methodology, and more insight can be found in Appendix F.

1. Coryn et al. (2009) provides details on a modified SCM used in NGO program IA.
2. Olson et al. (2011) provides details on a modified SCM used to evaluate medical education program's impact on implementing tobacco cessation programs.
3. Sim et al. (2019) provides details on a modified SCM used to evaluate psychosocial capacity building following earthquakes in China.
4. Medina et al. (2015) provides details on a modified SCM used to evaluate a medical education program's impact on research quality.
5. Bailey et al. (2006) provides details on a modified SCM used to evaluate a health and housing program.
6. Brinkerhoff (2005) provides the foundations of the original methodology design.

**Box 1:** *Important readings on SCM and previous modifications*

If the NS believe that this methodology is worth pursuing after learning the details, it will also be important to test the effectiveness at least one NS. This would entail training the FbF teams and the M&E teams on the methodology and using the next FbF intervention trigger to test this impact assessment technique. A lesson's learned study would need to be done on the impact assessment, and there should be feedback from the NS. As one interviewee said, the NS should be leading the process and have plenty of input. They are the ones that will be conducting these studies and should be comfortable with it. Additionally, funders and other stakeholders should have an input, as they will need the information from these impact assessments to make decisions on future funding and program direction.

## **6. Conclusion**

Gathering high quality data on impacts has been difficult in both humanitarian and A-A interventions. The most common methodology used, quasi-experimental studies, have proven challenging in terms of capacity and resources for many NS to conduct. The SCM methodology, with modifications to enhance the rigor and provide a comparison group, could potentially provide a solution to some of the methodological issues that have limited NS impact assessments in the past. Understanding the impact of different FbF interventions can help improve the interventions themselves as well as provide an evidence base that can be used to justify future FbF intervention developments. While this methodology has potential, it is completely untested in this context. More research and consultation are needed from different NS to determine whether this methodology will be suitable going forward. A pilot study using this methodology should be done to determine the limitations and benefits of the methodology in a real-life scenario.



## 7. References

- Anguko, A. (2019). Process Tracing as a methodology for evaluating small sample size interventions. *eVALUation Matters, Second Quarter 2019*, 19-27.  
<https://doi.org/10.13140/RG.2.2.31211.28963>
- Bailey, R., Clinton, J., Laing, P., Nunns, H., & Roorda, M. (2006, September 6-9). *Success Case Method: Uses and adaptation in New Zealand housing evaluation* [Conference presentation]. Australia Evaluation Society International Conference, Darwin, Australia. <https://aes.asn.au/images/images-old/stories/files/conferences/2006/papers/037%20Janet%20Clinton.pdf>
- Balbach, E. D. (1999). *Using case studies to do program evaluation*. California Department of Health Services.  
<https://www.betterevaluation.org/sites/default/files/ProgramEvaluation.pdf>
- Bamberger, M., Rugh, J., Church, M., & Fort, L. (2004). Shoestring evaluation: Designing impact evaluations under budget, time, and data constraints. *The American Journal of Evaluation*, 25(1).  
<https://doi.org/10.1177/109821400402500102>
- Bell, C. & McDonald, D. (2006, September 6-9). *The success case method: A simple evaluation tool to identify critical success factors and program impact* [Conference presentation]. Australia Evaluation Society International Conference, Darwin, Australia. <https://aes.asn.au/images/images-old/stories/files/conferences/2006/papers/063%20Catherine%20Bell.pdf>
- Bousquet, J., Kiemtore, I., Mvukiyehe, E. & Verpoorten, M. (2021). *Strengthening resilience in fragile and conflict-affected societies: evidence from a randomized control trial of a CDD program with a conflict resolution dimension in Eastern Congo*. The World Bank. <https://doi.org/10.1257/rct.7602-1.0>.
- Brinkerhoff, R. O. (2005). The Success Case Method: A strategic evaluation approach to increase the value and effect of training. *Advances in Developing Human Resources*, 7(86). <https://doi.org/10.1177/1523422304272172>

- Buttenheim, A. (2010). Impact evaluation in the post-disaster setting: a case study of the 2005 Pakistan earthquake. *Journal of Development Effectiveness*, 2(2), 197-227. <https://doi.org/10.1080/19439342.2010.487942>
- Canto-Blundo, G., Laderach, P., Waldock, J., & Camacho, K. (2017). Learning through monitoring, evaluation and adaptations of the “Outcome Harvesting” tool. *Cahiers Agricultures*, 26. <https://doi.org/10.1051/cagri/2017054>
- Catley, A., Burns, J., Abebe, D., & Suji, O. (2013). *Participatory Impact Assessment: A Design Guide*. Feinstein International Center, Tufts University, Somerville. <https://fic.tufts.edu/publication-item/participatory-impact-assessment-a-design-guide/>
- Casey, S. (2015). Evaluations of reproductive health programs in humanitarian settings: a systematic review. *Conflict and Health*, 9(S1). Available from: <http://www.conflictandhealth.com/content/9/S1/S1>
- Chapman, J. & Mancini, A. (2008). *Impact assessment: Drivers, dilemmas, and deliberations*. Sightsavers International
- Clinton, J., Mahony, F., Broadbent, R., Bullen, C. & Kerns, R. (2007). *The Healthy Housing Programme: Report of the outcome evaluation (year three)*. Housing New Zealand Corporation. [https://thehub.swa.govt.nz/assets/documents/Healthy%20Housing%20Evaluation%20\(Year%20Three\)%20July%202007.pdf](https://thehub.swa.govt.nz/assets/documents/Healthy%20Housing%20Evaluation%20(Year%20Three)%20July%202007.pdf)
- Copestake, J. (2014). Credible impact evaluations in complex contexts: confirmatory and exploratory approaches. *Evaluation*, 20(4), 412-427. <https://doi.org/10.1177/1356389014550559>
- Copestake, J., Allan, C., van Bekkum, W., Belay, M., Goshu, T., Mvula, P., Remnant, F., Thomas, E., & Zerahun, Z. (2018). Managing relationships in qualitative impact evaluation of international development: QUIP choreography as a case study. *Evaluation*, 24(2), 169-184. <https://doi.org/10.1177/1356389018763243>
- Coryn, C. L. S., Schroter, D. C. & Hanssen, C. E. (2009). Adding a time-series design element to the Success Case Method to improve methodological rigor.

- American Journal of Evaluation*, 30(1), 80-92.  
<https://doi.org/10.1177/1098214008326557>
- Coughlan de Perez, E., van der Hurk, B., van Aalst, M. K., Jongman, B., Klose, T., & Suarez, P. (2015). Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Natural Hazards and Earth System Sciences*, 15, 895-904. <https://doi.org/10.5194/nhess-15-895-2015>
- Dart, J. & Roberts, M. (2014) *Collaborative outcomes reporting*. BetterEvaluation. Available at: <http://betterevaluation.org/plan/approach/cort>
- Davies, R., & Dart, J. (2005). *The 'Most Significant Change' (MSC) technique: A guide to its use*. Care International, London, UK.  
<https://doi.org/10.13140/RG.2.1.4305.3606>
- DFID. (2012). *Broadening the range of designs and methods for impact evaluations, (Working Paper 38)*. Department for International Development.  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/67427/design-method-impact-eval.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf)
- Dickson, K. & Bangpan, M. What are the barriers to, and facilitators of, implementing and receiving MHPSS programmes delivered to populations affected by humanitarian emergencies? A qualitative evidence synthesis. *Global Mental Health* (2018), 5, e21. <https://doi.org/10.1017/gmh.2018.12>
- Duflo, E. & Kremer, M. (2003, July 15-16). *Use of randomization in the evaluation of development effectiveness* [Conference presentation]. World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness, Washington, D.C. <https://economics.mit.edu/files/765>
- Doocy, S. & Tappis, H. (2017). *Cash-based approaches in humanitarian emergencies: a systematic review*. Campbell Systematic Reviews.  
<https://doi.org/10.4073/csr.2017.17>
- Duvendack, M., Hombrados, J. C., Palmer-Jones, R. & Waddington, H. (2012). Assessing 'what works' in international development: meta-analysis for sophisticated dummies. *Journal of Development Effectiveness*, 4(3), 456-471.  
<https://doi.org/10.1080/19439342.2012.710642>

- Duvendack, M., Palmer-Jones, R. & Vaessen, J. (2014). Meta-analysis of the impact of microcredit on women's control over household decisions: methodological issues and substantive findings. *Journal of Development Effectiveness*, 6(2), 73-96, <https://doi.org/10.1080/19439342.2014.903289>
- FAO. (2018). *Impact of early warning early action: Protecting pastoralist livelihoods ahead of drought*. Food and Agriculture Organization of the United Nations: Rome. <http://www.fao.org/publications/card/en/c/CA0227EN/>
- FAO. (2018a). *Impact of early warning early action: Protecting the livelihoods of herders from dzud winter*. Food and Agriculture Organization of the United Nations: Rome. <http://www.fao.org/3/ca2181en/CA2181EN.pdf>
- FAO. (2019). *Impact of early warning early action: Protecting agropastoralist livelihoods ahead of drought*. Food and Agriculture Organization of the United Nations: Rome. <http://www.fao.org/documents/card/en/c/CA4653EN/>
- FAO. (2020). *Impact of early warning early action: Exploring the interplay between El Niño-induced drought, conflict, and gender*. Food and Agriculture Organization of the United Nations: Rome. <http://www.fao.org/documents/card/en/c/CA9371EN/>
- Ferrah, S., Gazeaud, J., Khan, N., Mvukiyehe, E., Shrotri, V., Streck, O. & Zineb, S. B. (2021). *Enhancing female entrepreneurship through cash grants: Experimental evidence from rural Tunisia*. World Bank's Middle East and North Africa Gender Innovation Lab. <http://hdl.handle.net/10986/35863>
- Fuller, R. & Lain, J. (2017). *Building resilience: A meta-analysis of Oxfam's resilience Effective Reviews*. Oxfam. <https://policy-practice.oxfam.org/resources/building-resilience-a-meta-analysis-of-oxfams-resilience-effectiveness-reviews-620356/>
- Garbero, A. & Chichaibelu, B. B. (2018). *Impact assessment report: The agricultural sector development programme-livestock and the agriculture service support programme, Tanzania*. IFAD: Rome. <https://www.ifad.org/en/web/knowledge/-/publication/impact-assessment-agricultural-sector-development-programme-livestock-asdp-l-and-agriculture-service-support-programme-assp->

- German Red Cross. (2017). *Forecast-based Financing: A policy overview*. German Red Cross: Berlin.  
[https://www.drk.de/fileadmin/user\\_upload/FBF/Policy\\_overview\\_sep\\_2017.pdf](https://www.drk.de/fileadmin/user_upload/FBF/Policy_overview_sep_2017.pdf)
- German Red Cross. (2019). *Forecast-based financing: A new era for the humanitarian system*. German Red Cross: Berlin. [https://www.forecast-based-financing.org/wp-content/uploads/2019/03/DRK\\_Broschuere\\_2019\\_new\\_era.pdf](https://www.forecast-based-financing.org/wp-content/uploads/2019/03/DRK_Broschuere_2019_new_era.pdf)
- German Red Cross. (2020). *Our Projects*. <https://www.forecast-based-financing.org/our-projects/>
- Gertler, P. J., Martinez, S., Premand, P., Rawling, L. B., Vermeersch, C. M. J. (2011). *Impact evaluation in practice*. The World Bank.  
<https://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice>
- Gros, C., Bailey, M., Schwager, S., Hassan, A., Zingg, R., Uddin, M. M, Shahjahan, M., Islam, H., Lux, S., Jaime, C., Coughlan De Perez, E. (2019). Household-level effects of providing forecast-based cash in anticipation of extreme weather events: Quasi-experimental evidence from humanitarian interventions in the 2017 floods in Bangladesh. *International Journal of Disaster Risk Reduction*, 41.  
<https://doi.org/10.1016/j.ijdrr.2019.101275>
- Gros, C., Easton-Calabria, E., Bailey, M., DAgys, K., Coughlan de Perez, E., Sharavnyambuu, M., Kruczkiewicz, A. (2020). The effectiveness of forecast-based humanitarian assistance in anticipation of extreme winters: Evidence from an intervention for vulnerable herders in Mongolia. *Disasters* (accepted article).  
<https://doi.org/10.1111/disa.12467>
- Hofmann, C. A., Roberts, L., Shoham, J. & Harvey, P. (2004). *Measuring the impact of humanitarian aid: a review of current practices*. Overseas Development Institute. <https://odi.org/en/about/our-work/measuring-the-impact-of-humanitarian-aid-a-review-of-current-practice/>
- Holvand, I. (2005). *Successful communication: A toolkit for researchers and civil society organizations (Working Paper 227)*. Overseas Development Institute: London <https://cdn.odi.org/media/documents/192.pdf>

- Hulme, D. (2000). Impact assessment methodologies for microfinance: theory, experience, and better practice. *World Development*, 28(1), 79-98.  
[https://doi.org/10.1016/S0305-750X\(99\)00119-9](https://doi.org/10.1016/S0305-750X(99)00119-9)
- Israel, G. D. (1992). *Determining sample size*. University of Florida Cooperative Extension Service, Institute of Food and Agriculture Sciences, EDIS, Florida.
- Jjemba, E. W., Mwebaze, B. K., Arrighi, J., Coughlan de Perez, E., & Bailey, M. (2018). Chapter 19-Forecast-based financing and climate change adaptation: Uganda makes history using science to prepare for floods. In Z. Zommers & K. Alverson (Eds.), *Resilience: The Science of Adaptation to Climate Change* (pp. 237-242). Elsevier. <https://doi.org/10.1016/C2016-0-02121-6>
- Jones, B. (1998). *Beneficiary Assessment for monitoring: the Zambia social recovery project*. World Bank, Social Development Notes, Note No. 36
- Jones, N., Jones, H., Steer, L. & Datta, A. (2009). *Improving impact evaluation production and use (working paper 300)*. Overseas Development Institute.  
<https://www.files.ethz.ch/isn/98153/wp300.pdf>
- Kendall, P. C., Comer, J. S., & Chow, C. (2013). The randomized controlled trial: Basics and beyond. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 40–61). Oxford University Press.  
<https://doi.org/10.1093/oxfordhb/9780199793549.013.0004>
- Leeuw, F. & Vaessen, J. (2009). *Impact evaluation and development: NONIE guidance on impact evaluation*. Network of Networks on Impact Evaluation.  
<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/411821468313779505/impact-evaluations-and-development-nonie-guidance-on-impact-evaluation>
- Malek, M. A., Ahasan, A., Hossain, M. A., Ahmed, M. S., Hossain, M., and Reza, M. H. (2015). *Impact assessment of credit programme for the tenant farmers in Bangladesh, 3ie Grantee Final Report*. New Delhi: International Initiative for Impact Evaluation (3ie). <https://www.3ieimpact.org/sites/default/files/2018-08/Impact%20assessment%20of%20credit%20programme%20for%20the%20tenant%20farmers%20in%20Bangladesh.pdf>

- Malterud, K., Siersma, V. D., & Gaussora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753-1760. <https://doi.org/10.1177/1049732315617444>
- Marston, C., Renedo, A., McGowan, C. R., & Portela, A. (2013). Effects of community participation on improving uptake of skilled care for maternal and newborn health: A systematic review. *PLoS ONE* 8(2): <https://doi.org/10.1371/journal.pone.0055012>
- Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research? A review of qualitative interviews in is research. *Journal of Computer Information Systems*, 54(1), 11-22. <https://doi.org/10.1080/08874417.2013.11645667>
- Mayne, J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *Canadian Journal of Program Evaluation*, 16, 1-24. [https://www.researchgate.net/publication/279533461\\_Addressing\\_Attribution\\_Through\\_Contribution\\_Analysis\\_Using\\_Performance\\_Measures\\_Sensibly](https://www.researchgate.net/publication/279533461_Addressing_Attribution_Through_Contribution_Analysis_Using_Performance_Measures_Sensibly)
- Mayne, J. (2008). *Contribution analysis: An approach to exploring cause and effect (Brief 16)*. The Institutional Learning and Change Initiative. [https://www.researchgate.net/publication/46472564\\_Contribution\\_analysis\\_An\\_approach\\_to\\_exploring\\_cause\\_and\\_effect](https://www.researchgate.net/publication/46472564_Contribution_analysis_An_approach_to_exploring_cause_and_effect)
- Medina, L., Acosta-Perez, E., Velez, C., Martinez, G., Rivera, M., Sardinias, L. & Pattatucci, A. (2015). Training and capacity building evaluation: Maximizing resources and results with Case Success Method. *Evaluation Program Planning*, 52, 126-132. <https://doi.org/10.1016/j.evalprogplan.2015.03.008>
- Neubert, S. (2010). *MAPP: a participatory method for impact assessment of programs and projects*. German Development Institute. <https://zewo.ch/wp-content/uploads/2019/07/MAPP-Description.pdf>



- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16, 1-13. <https://doi.org/10.1177/1609406917733847>
- Olson, C. A., Shershneva, M. B. & Brownstein, M. H. (2011). Peering inside the clock: Using Success Case Method to determine how and why practice-based educational interventions succeed. *Journal of Continuing Education in the Health Professions*, 31(S1), S50-S59. <https://doi.org/10.1002/chp.20148>
- OXFAM. (2019). *Understanding survey sampling*. OXFAM International Research Guidelines. Available at: [www.oxfam.org.uk/policyandpractice](http://www.oxfam.org.uk/policyandpractice)
- Piggott-Irvine, E., Marshall, S. & Aitken, H. (2009). *The three R's of the success case model – Recruitment, response, rigour*. Proceedings of the Australasian Evaluation Society International Conference. Retrieved from <http://www.aes.asn.au/conferences/2009/Papers/Piggott-Irvine, Marshall & Aitken.pdf>
- Proudlock, K., Ramalingham, B. & Sandison, P. (2020). *Improving humanitarian impact assessment: bridging theory and practice*. ALNAP Review of Humanitarian Action. <https://www.alnap.org/system/files/content/resource/files/main/8rhach2.pdf>
- Puri, J., Aladysheva, A., Iversen, V., Ghorpade, Y., & Bruck, T. (2014). *What methods may be used in impact evaluations of humanitarian assistance? (Working paper 22)*. International Initiative for Impact Evaluation. [https://www.3ieimpact.org/sites/default/files/2019-01/wp\\_22\\_humanitarian\\_methods\\_working\\_paper-top\\_0.pdf](https://www.3ieimpact.org/sites/default/files/2019-01/wp_22_humanitarian_methods_working_paper-top_0.pdf)
- Ramesh, A., Blanchet, K., Ensink, J. H. J., & Roberts, B. (2015). Evidence on the effectiveness of water, sanitation, and hygiene (WASH) interventions on health outcomes in humanitarian crises: A systematic review. *PLoS ONE*, 10(9). <https://doi.org/10.1371/journal.pone.0124688>
- Ravindra, A. (2004). *An assessment of impact of Bangalore citizen report card on the performance of public agencies (ECD working paper series 12)*. The World Bank Operations Evaluation Department.



[http://lnweb18.worldbank.org/oed/oeddolib.nsf/24cc3bb1f94ae11c85256808006a0046/d241684df81fce2785256ead0062de10/\\$FILE/ecd\\_wp\\_12.pdf](http://lnweb18.worldbank.org/oed/oeddolib.nsf/24cc3bb1f94ae11c85256808006a0046/d241684df81fce2785256ead0062de10/$FILE/ecd_wp_12.pdf)

Roche, C. (2000). Impact Assessment: Seeing the Wood and the Trees. *Development in Practice*, 10(3). <https://www.jstor.org/stable/4029582>

Roughley, A., & Dart, J. (2009). *Developing a performance story report: User guide*. Commonwealth of Australia. <https://www.cifor.org/wp-content/uploads/dfid/KNOWFOR%20monitoring%20tool%20kit/9.%20PSR%20guide.pdf>

Salmen, L. F. (1999). *Beneficiary assessment manual for social funds*. Environmentally and Socially Sustainable Development Network. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/951471468739336407/beneficiary-assessment-manual-for-social-funds>

Salmen, L. F. (2002). *Beneficiary Assessment: An approach described*. Environment Department Papers, Social Assessment Series, Paper No. 010. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/802501468739312293/beneficiary-assessment-an-approach-described>

Sango, D. & Dube, S. D. (2014). *Most Significant Change stories: Zimbabwe emergency drought mitigation operation for livestock project*. Smart Development Works.

Schatz, F. & Welle, K. (2016). *Qualitative Comparative Analysis: A valuable approach to add to the evaluator's toolbox? Lessons from recent applications*. CDI Practice Paper No. 13. <http://archive.ids.ac.uk/cdi/publications/qualitative-comparative-analysis-valuable-approach-add-evaluator%E2%80%99s-%E2%80%99toolbox%E2%80%99lessons.html>

Serratt, O. (2009). *The Most Significant Change technique*. Asian Development Bank. <https://www.adb.org/sites/default/files/publication/27613/most-significant-change.pdf>

- Shadish, W. R, Cook. T. D., & Campbell, D. T. (2001). *Modern descriptions of experiments*. In W. R. Shadish, W. R. Cook, & D. T. Campbell (Eds.) *Experimental and quasi-experimental designs for generalized causal inference* (pp. 14-18). Boston: Houghton Mifflin Company.
- Singh, A. S. & Masuku, M. B. (2014). Sampling techniques & determination of sample size in applied statistics research: an overview. *International Journal of Economics, Commerce and Management*, 2(11), 1-22. Available at: [https://www.researchgate.net/publication/341552596\\_Sampling\\_Techniques\\_and\\_Determination\\_of\\_Sample\\_Size\\_in\\_Applied\\_Statistics\\_Research\\_An\\_Overview](https://www.researchgate.net/publication/341552596_Sampling_Techniques_and_Determination_of_Sample_Size_in_Applied_Statistics_Research_An_Overview)
- Sim, T., Lau, J., Cui, K., & Wei, H. (2019). Post-disaster psychosocial capacity building for women in a Chinese rural village. *International Journal of Disaster Risk Science*, 10, 193-203. <https://doi.org/10.1007/s13753-019-0221-1>
- Simister, N. & Scholz, V. (2017). Qualitative Comparative Analysis. INTRAC. <https://www.intrac.org/wpcms/wp-content/uploads/2017/01/Qualitative-comparative-analysis.pdf>
- SMART. (2012). *Sampling methods and sample size calculations for the SMART Methodology*. SMART. <https://www.alnap.org/help-library/sampling-methods-and-sample-size-calculation-for-the-smart-methodology>
- Tanner, T., Gray, B., Guigma, K., Iqbal, J., Levine, S., MacLeod, D., Nahar, K., Rejve, K., & Venton, C. C. (2019). *Scaling up early action: Lessons, challenges, and future potential in Bangladesh (Working Paper 547)*. Overseas Development Institute. <https://cdn.odi.org/media/documents/12641.pdf>
- The World Bank. (2004). *Citizen report card surveys: A note on the concept and methodology (Note No. 91)*. World Bank Social Development Notes. <https://openknowledge.worldbank.org/handle/10986/11277>
- USAID. (2021). *Technical guidance for monitoring, evaluation, and reporting for emergency activities*. USAID Bureau for Humanitarian Assistance. Available at: <https://www.usaid.gov/bha-guidelines/draft-bha-emergency-me-guidance>
- Watson, C. (2008). *Impact assessment of humanitarian response: A review of the literature*. Fienstein International Center.

<https://www.alnap.org/system/files/content/resource/files/main/impact-10-07-08.pdf>

Weingärtner, L., Pforr, T. & Wilkinson, E. (n.d). *The evidence base on anticipatory action*. World Food Program. <https://www.wfp.org/publications/evidence-base-anticipatory-action>

Weingärtner, L. & Wilkinson, E. (2019). *Anticipatory crisis financing and action: concepts, initiative, and evidence*. Overseas Development Institute. <https://odi.org/en/publications/anticipatory-crisis-financing-and-action-concepts-initiatives-and-evidence/>

Westthorp, G. (2014). *Realist impact evaluation: An introduction*. Overseas Development Institute. <https://cdn.odi.org/media/documents/9138.pdf>

WFP. (2009). *Emergency food security assessment handbook-second edition*. World Food Program, Food Security Analysis Program. <https://www.wfp.org/publications/emergency-food-security-assessment-handbook>

White, H. & Phillips, D. (2012). *Addressing Attribution of Cause and Effect in Small-n Impact Evaluations: Towards an integrated framework*. Working Paper 15, International Initiative for Impact Evaluation. [https://www.3ieimpact.org/sites/default/files/2019-01/working\\_paper\\_15.pdf](https://www.3ieimpact.org/sites/default/files/2019-01/working_paper_15.pdf)

White, H., & Raitzer, D. A. (2017). *Impact evaluation of development intervention: a practical guide*. Asian Development Bank. <https://www.adb.org/publications/impact-evaluation-development-interventions-practical-guide>

White, H., & Sabarwal, S. (2014). *Quasi-experimental design and methods (Methodological briefs impact evaluation No. 8)*. UNICEF Office of Research. <https://www.unicef-irc.org/publications/753-quasi-experimental-design-and-methods-methodological-briefs-impact-evaluation-no.html>

## **Appendix A: Semi-Structured Interview Guides**

### **Impact Assessment on a Shoestring: Determining Impacts of Forecast Based Financing Programs in Limited Resource Settings**

#### **Scoping Phase Questionnaire-National Society Members**

##### **Introduction:**

Thank you for taking the time to meet with me today. I am conducting this interview to learn more about what is currently being done at the sites to monitor the programs and to learn your thoughts on what could be done differently. I would like to ask you some questions about your experience with your FbF programs, your ideas and thoughts on impact assessments, previous impact assessment and monitoring experiences, and site capacity and resources that are available to conduct assessments. The interview should take no more than 45 minutes. If it is ok with you, I will be recording the conversation because I don't want to miss any of your comments. I may also take some notes. All your responses will be kept confidential. This means that your interview responses will only be shared with research team members and we will ensure that any information we include in our report does not identify you as the respondent. Remember, you don't have to talk about anything you don't want to and you may end the interview at any time. We appreciate your answering these questions as honestly as possible. Please feel free to ask me if you have any questions.

##### **Section 1: Outcomes of Interest**

- What you most interested in learning about the impacts your FbF program?
- What do you believe are the most important outcomes to assess when looking at FbF impacts? (Probe is ensure that interviewees refer to their EAP, which describe what the outcomes are meant to achieve)
  - o Health and well-being (mitigating loss of life, injury, disease, mental health issues, ect.)
  - o Assets (house, livestock, documents, etc)
  - o Household finances?
  - o Evacuations?
  - o Food security?
  - o Other?
- What information do you believe is most practical to collect/measure with the resources available at your site?
- What do you need to collect for reporting requirements?

##### **Section 2: Current Practices**

- Has your FbF Early Action Protocol (EAP) ever been activated based on a trigger?
- (If not:) Has your FbF project/programme done a practical field simulation of the anticipatory actions, involving potential beneficiary households or communities?

- Have you ever conducted or commissioned an impact assessment of your FbF EAP before? By impact assessment I mean an analysis of the effects of the anticipatory actions on the beneficiaries in relation to what the EAP is meant to achieve?
  - o If yes, what was the experience like?
    - What type of assessment was done?
    - What outcome variables were looked at?
    - What resources were required?
      - How much did this cost in total?
      - Did you have to hire any external support, for example: enumerators, sample design, or a data analysis consultant?
  - o If no, do you have an impact assessment plan for your EAP in place?
    - What methods are involved?
    - Are current staff trained to conduct these assessments?
- What barriers do you perceive to conducting impact assessments?

### **Section 3: Capacity and Resources**

- What type of M&E do you, or the National Society, normally do for your programmes or projects? (prompt for typical categories, such as: post distribution (output) monitoring, beneficiary satisfaction surveys, after action reviews or lessons learned workshops, narrative donor reports, qualitative evaluations)
- Would you say there is interest in conducting reliable impact assessments of the anticipatory action EAP, in case it is triggered – noting that EAP funding from the DREF comes with a dedicated budget line for M&E?
- What resources do you have available to conduct impact assessments of your EAP? By impact assessment I mean determining the impact of activation on beneficiaries
  - o Do you have a dedicated M&E focal point?
    - Do they have the capacity to take on assessment work on top of their current activities?
  - o If you wanted to do a quantitative survey involving statistical analysis (for example, testing for statistically significant differences between groups), Is the M&E focal point a trained statistician, or someone else on the team?
  - o Are there any existing partnerships with universities/external partners that could assist in assessment activities?
    - If yes, what is their level of involvement with the site?
      - Is there more than one partner involved?
  - o How many members of staff work on FbF activities? Will members of staff be available to conduct interviews or do monitoring during an assessment?
  - o Is anyone on the staff trained in conducting interviews or surveys?
- What resources do you anticipate needing to conduct impact assessments?
- Is the 'FbA by the DREF' funding for your anticipatory action M&E sufficient to conduct an impact assessment or evaluation in case of an activation? What else would be needed?
- How quickly after an EAP activation can an impact assessment begin with the resources currently available?

# **Impact Assessment on a Shoestring: Determining Impacts of Forecast Based Financing Programs in Limited Resource Settings**

## **Scoping Phase Questionnaire-Other Key Informants**

### **Introduction:**

Thank you for taking the time to meet with me today. I am conducting this interview to learn more about what is currently being done to monitor FbF programs and to learn your thoughts on what could be done differently. I would like to ask you some questions about your experience with your FbF programs, your ideas and thoughts on impact assessments, and previous impact assessment and monitoring experiences. When I say 'impact assessment', I mean attempts to measure whether and to what extent FbF anticipatory actions make a difference to the individual beneficiary, household or community. The interview should take about 30-45 minutes. If it is ok with you, I will be recording the conversation because I don't want to miss any of your comments. I may also take some notes. All your responses will be kept confidential. This means that your interview responses will only be shared with research team members and we will ensure that any information we include in our report does not identify you as the respondent. Remember, you don't have to talk about anything you don't want to and you may end the interview at any time. We appreciate your answering these questions as honestly as possible. Please feel free to ask me if you have any questions.

### **Section 1: General**

1. What do we know about the effectiveness of FbF programs in general?
  - a. Are there any previous pieces of evidence that stand out to you?
  - b. Are there things that you would like to know that have not been studied yet?
2. What do you know about the current evidence landscape?
  - a. What is working well?
  - b. What is currently missing?
3. In your opinion, why are things missing in the evidence landscape?
  - a. lack of guidance?
  - b. M&E capacity of FbF implementers?
  - c. Learning interests?

### **Section 2: Evaluation changes**

1. In your opinion, where should we go from here in terms of impact assessment?
  - a. What changes would you like to see?
  - b. Are these changes practical given the resources available?
2. What type of evidence is most important for improving the design and implementation of FbF interventions?
3. What information is the most practical to collect?
4. For your position, or from your perspective, what types of questions are most important to ask?
  - a. Are there specific outcomes of interest that you would like to see investigated?

- b. What questions would be the most helpful for your position?
5. What do you believe are the main constraints to collecting the information?
6. What skills to you think are required for these assessments?
7. Do you have any ideas for new methods of FbF assessments?

## NVivo Code Book

Name	Description
<b>Capacity</b>	national society capacity available to do IA
<b>Community ties</b>	Relationship with the community
<b>M&amp;E team</b>	Monitoring and evaluation teams at NS
<b>questionnaire design</b>	Capacity to design questionnaires for IA
<b>staff</b>	Other staff available besides M&E and statistics personnel
<b>Statistics</b>	Statistical ability for sampling and analysis
<b>Technology</b>	Technology used and available for IA (cell phones)
<b>Volunteers</b>	Information about volunteers being used to collect data for IA and other data collection needs
<b>Challenges</b>	Challenges to the current design for IA
<b>Current Practices</b>	Current practices in IA at different FbF programs
<b>Important Information</b>	What information programs need or want to gather about their FbF programs
<b>best practices review</b>	Some sites have or want a best practices review
<b>outcomes of interest</b>	Outcomes of interest for the intervention. What they are and what should be collected
<b>qualitative information</b>	Qualitative data to be collect, what kind and why
<b>reporting requirements</b>	Some information is required by the IFRC or the funders

Name	Description
<b>New ways to gather data</b>	Novel ideas for different ways to collect impact data that does not include survey administration
<b>Cell phone data</b>	Collecting data via cell phones or collecting cell phone data
<b>remote sensing</b>	Remote sensing or satellite data
<b>What they want in terms of IA</b>	Ideas of what impact assessments should be about and what they should look at



## Appendix B: Methodology Scoring Guide

### Scoring system for methodologies:

	1	2	3	4	5
<b>Capacity Required*</b>	High capacity required to conduct methodology- this could be strong statistical ability needed or strong knowledge of sampling skills	Moderate to high capacity required- moderate stat skills or capacity in planning phase	Moderate capacity required	Low to moderate capacity required	Low capacity required- only qualitative data analysis
<b>Preference of National Societies**</b>	Something that the NS said they found difficult- finding some of the sampled people	Meeting few preferences	Neutral preference (either no mention or mentioned that it was fine)	Meeting most, but not all preferences	Meeting all preferences that the NS said they wanted to include (example: qualitative research, comparison, surveys)
<b>Complexity of Design/administration†</b>	Highly complex, either in design or administration of methodology	Moderate to high complexity	Moderate complexity	Moderate to low complexity	Low complexity
<b>Cost ‡</b>	High cost of administration	Moderate to high cost of administration	Moderate cost	Moderate to low cost	Low cost

\*Capacity can be capacity in data analysis, sampling design, pre-activity planning or data trolling, questionnaire development, etc.

\*\* Preference comes from interviews of NS employees or other stakeholders. They said things like wanting to interview non-beneficiaries, qualitative interviews, surveys being easy

† Complexity of design means that there are many different steps required or there is something complicated in the design. Complexity of administration means that something about administering the methodology is complex. This includes things like large number of revisiting, very large sample sizes, ect.

‡ Cost comes from many different things: hiring statisticians, paying volunteers to conduct a lot of surveys, hiring cars to reach many different participants, putting together workshops with many stakeholders.

## **Appendix C: Detailed Information on Methodologies Reviewed**

### **Quantitative:**

#### **1. Randomized Control Trial**

The basics of randomized control trials is that participants are randomized into a treatment (or in this case beneficiary) group and a control (or non-beneficiary) group. It is considered the gold standard in study design, as it eliminates a lot of the bias that exists in other methodologies. If randomization is done correctly, there is more confidence that the intervention has led to the outcome, because both external and internal validity can be controlled for. It rules out the possibility that other factors have led to the outcome. There is a broad history of use of this methodology in the development field.

##### **1. Pre-data collection activities**

The first step of an RCT methodology is to design the intervention, as participants will have to be randomized to either receive the intervention or not. This means that all the specifics of the intervention, plus theories and hypotheses of what the effect will be, must be determined first.

There are several other activities that need to be completed before data collection can begin, the most important being randomization and sampling. Different studies will do sampling in different ways, depending on the size of the intervention. If the intervention takes place over a large geographical area, or if the intervention size is very large, then cluster sampling might be employed. For RCTs, sampling happens before the intervention is given, because of randomization, while most other methodologies will sample after interventions have been given. There are also different options for where to randomize. Many studies have randomized interventions at a district or county level, while others choose to randomize individuals. This will depend on the exact design.

This is a very systematic process that involves statisticians. Sample sizes are determined based on power calculations and the effect size. This will require the capacity to understand and conduct sample size calculations. These designs usually have quite large sample sizes, with one study having over 1000 participants in each group. Other activities such as training volunteers and structured questionnaire development also take place at this stage.

##### **2. Data Collection**

For development programs, which are usually longer term than humanitarian interventions, RCTs involve collecting baseline and endline data from both treatment and control groups. Information is usually gathered on sociodemographic characteristics, intervention uptake, changes in outcome conditions, and any other information that is needed to assess the impact. In RCTs, this information will need to be collected at the start of the study, ideal before the intervention has started. This will give a good understanding of the baseline condition. Someone on staff must be responsible for managing the data that is collected and ensuring the quality. The number of enumerators will depend on the resources available, but more enumerators will mean that collection happens more rapidly.

##### **3. Data Analysis**

There are two stages of analysis in most RCTs, the baseline analysis and then the larger analysis with endline data. The baseline analysis is usually meant to confirm that there are no statistical differences in variables between the control and treatment group. Once this is established, then any differences seen at the end should be contributed to the intervention. This typically involves simple t-tests and normalized differences.

Endline data analysis is more complicated, with knowledge of more complex statistics required. Multi-variable regression, other regression techniques, and potentially disaggregating results to look at sub-groups. Analysis of this data requires someone with a strong knowledge of statistics and analytical tools.

#### 4. Evidence Gathered

If done correctly, this methodology will provide numerical evidence that the intervention led to the impacts being measured. It will also provide numeric baseline data and numeric data on the different outcomes of the intervention.

Sources:

Kendall, P. C., Comer, J. S., & Chow, C. (2013). The randomized controlled trial: Basics and beyond. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 40–61). Oxford University Press.

Malek, M. A., Ahasan, A., Hossain, M. A., Ahmed, M. S., Hossain, M., and Reza, M. H. (2015). *Impact assessment of credit programme for the tenant farmers in Bangladesh, 3ie Grantee Final Report*. New Delhi: International Initiative for Impact Evaluation (3ie)

Duflo, E. & Kremer, M. (2003, July 15-16). Use of randomization in the evaluation of development effectiveness [paper presentation]. World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness, Washington, D.C.

Puri, J., Aladysheva, A., Iversen, V., Ghorpade, Y., & Bruck, T. (2014). What methods may be used in impact evaluations of humanitarian assistance? (Working paper 22). International Initiative for Impact Evaluation.

## 2. Quasi-experimental

Quasi-experimental studies are an adaptation of experimental studies such as RCT. They provide similar control and intervention groups, but without randomizations at the beginning of the intervention. The goal is to try and establish a counterfactual, looking at what would have happened if the intervention had not occurred, without the randomization process. Typically, this design samples beneficiaries and non-beneficiaries after the intervention has taken place. During this sampling, it is necessary to find non-beneficiaries that closely match the demographic information of the beneficiaries, so that comparisons between the two can demonstrate the impacts of the program. This can be difficult in humanitarian situations where the interventions are meant to target the most vulnerable people.

#### 1. Pre-data collection activities

Prior to the collection of data, the comparison groups for beneficiaries need to be decided on. There are multiple methods for doing this matching. Most of these techniques involve gathering baseline data on non-beneficiaries. That data can come from data expressly gathered for the study or from background data collected for other purposes.

One of the methods for matching is propensity score matching, which was done in both Red Cross studies looked at in this report. Propensity score matching can be done with either baseline or endline data, but variables must be ones that have not been influenced by the intervention, so baseline data is more appropriate. The matching process involves several different statistical techniques, including logit regression and a matching algorithm, to compare values and match participants to non-participants. Once groups are determined, they are checked to see if there are any significant differences between the two groups. Ideally there should be no differences. For this and other matching techniques, strong statistical skills are needed. Other techniques including regression discontinuity design, ordinary least squares, and logistic regression can also be used.

Other things such as questionnaire design and sampling need to be done before data gathering. This technique requires large sample sizes.

## 2. Data Collection

Data is usually collected via large surveys. Volunteers or staff that are trained in administering these surveys can visit sampled participants to collect data. Data will need to be collected from both beneficiaries and non-beneficiaries. Depending on the design of the study and the operational situation, either baseline and endline or just endline data will need to be collected. The number of enumerators will depend on the resources available, but more enumerators will mean that collection happens more rapidly.

## 3. Data Analysis

There are several different ways to analyze the matched data. The most common is to compare the outcome of the intervention group to the outcomes in the control group to determine if there were significant differences. Other techniques include difference-in-difference, which compares changes in groups over time, and regression analyses. All of these techniques will require someone with a high level of statistical knowledge, as well as data that is well collected and cleaned.

## 4. Evidence Gathered

This methodology will provide numerical evidence on differences between those that received the interventions and those that did not. It also provides numerical point evidence on specific outcome indicators. Depending on the design, it can also provide numerical evidence on differences between different interventions or different levels of the same intervention. The evidence from this study design can make a strong case for attributing the intervention to the impact.

### Sources:

White, H., & Raitzer, D. A. (2017). *Impact evaluation of development intervention: a practical guide*. Asian Development Bank.

White, H., & Sabarwal, S. (2014). *Quasi-experimental design and methods (methodological briefs impact evaluation No. 8)*. UNICEF Office of Research.

Garbero, A. & Chichaibelu, B. B. (2018). *Impact assessment report: The agricultural sector development programme-livestock and the agriculture service support programme, Tanzania*. IFAD: Rome.

### **3. Case-Control Study**

Case-control studies originate from epidemiology. They are designed to match participants in the intervention group to those that have not had the intervention. This is similar to the matching done in quasi-experimental designs, where participants are matched on characteristics that do not affect the outcome. The researcher needs to determine the no unobserved characteristics are influencing the outcomes.

#### **1. Pre-data collection activities**

Prior to collecting data, sampling needs to occur. It is also essential that a bit of research is done on the matching variables and observations need to be made on what outcomes need to be looked at, as there may be several in FbF studies. Matching criteria needs to be established, as there are several variables that could be potentially used for matching. Over-matching will make selection difficult, while under-matching will result in comparisons that are not accurate.

#### **2. Data Collection**

Data needs to be collected in two stages. First, to match participants, data on matching variables needs to be collected. No other baseline data is needed. Once matching variables are established and matching is done, a larger data collection needs to take place. This will involve surveys that collect information on outcomes and factors that might have contributed to outcome. The number of enumerators will vary depending on resources, but fewer enumerators will lead to longer data collection times. This design uses smaller sample sizes than RCT, but large samples may still be required.

#### **3. Data Analysis**

There are several different ways to analyze the matched data. The most common is to compare the outcome of the intervention group to the outcomes in the control group to determine if there were significant differences.

#### **4. Evidence Gathered**

This technique can give good numeric evidence on differences between people that did and did not have a pre-determined outcome, which can be valuable information for the program. This can highlight some factors that will need to be taken into account in future similar interventions.

Sources:

Puri, J., Aladysheva, A., Iversen, V., Ghorpade, Y., & Bruck, T. (2014). What methods may be used in impact evaluations of humanitarian assistance? (Working paper 22). International Initiative for Impact Evaluation.

### **Qualitative:**

#### **1. Beneficiary Assessment**

Beneficiary Assessment is a systematic inquiry into people's values and beliefs regarding the intervention. It is used to improve the impact of interventions by gaining the views of beneficiaries. It is a methodology that is rarely used on it's own, it is most commonly used with quantitative methodologies to gain greater insight into interventions and people's

feelings about it. The theory behind the methodology is that beneficiaries usually lack a strong enough voice in the process. This methodology involves interviewing participants to better understand how they feel about interventions. Beneficiary Assessment is used by the World Bank to monitor their social programs.

## 1. pre-data collection activities

Prior to collecting data, several things need to be done. Whomever is the lead on the assessment will need to have a very strong understanding of the program, which requires becoming familiar with the socio-economic conditions of the areas as well as the institutional environment. This includes reading all program related documents and interviews with key personnel involved in developing and implementing the project. If time permits, an exploration of the area of the intervention, including talking to members of the community, is advised.

After gaining a real understanding of the context, the next step is to determine the objectives of the assessment. It is essential that the objectives are feasible and realistic in the context of the area. There are several typical objectives that are used in the methodology, including 'determining and examining factors underlying motivation to achieve intervention goals.' In the context of FbF programs, objectives could include 'understanding the underlying mechanisms behind positive outcomes'.

Selecting the interview team is also important. The number of people used will be determined by the sample size of participants. Team members will need to have very good interview skills. They can be people with university backgrounds, but as long as they have good communication and writing skills they can be used. Theoretically volunteers could be trained to do some interviews. Interviewers should be balanced in terms of gender, so that more open communication can take place.

Sampling also needs to be completed before data collection. Sample size needs to be large enough to get a good understanding of the feelings towards the intervention. Because of the intensive nature of the interviews, great understanding can be taken from a smaller number of participants. Purposive random sampling techniques are usually used, with stratification of the sample based on characteristics such as ethnicity, income, and gender. Typically 10-20% of communities are sampled, and in each community 5-20% of households are interviewed. Some designs also include interviewing non-beneficiaries with similar characteristics of beneficiary communities.

## 2. Data Collection

Interview questions include themes such as prioritization of needs in the community, wealth ranking, maintenance of programs, benefits or programs and any negative aspects. Trained volunteers or staff members conduct conversational interviews, usually in a interviewees home and in their local dialect. The objective is to gain in-depth information on beneficiary views. Focus groups can also take place, although some participants may not be as comfortable speaking in groups. Participant observation may also be used to gather information. This involves someone staying in the community for an extended period of time, and conducting case studies of 5-10 households, which are visited several times. Participant observation only occurs in a sub-set of studies because it is time consuming and more difficult.

### 3. Data Analysis

The final product of this methodology is a beneficiary assessment report. This will include analysis from the interviews. This will involve qualitative data analysis, where themes are taken from the interviews and synthesized to better understand how beneficiaries feel about the intervention, including if they feel there has been an impact.

### 4. Evidence Gathered

The evidence gathered is typically observations of the communities and changes that have been made due to the intervention. It also gives qualitative evidence on how the beneficiaries believe that the intervention helped them. This methodology is more geared towards long term development projects, but with a few small adaptations it could be used for FbF assessments.

Sources:

Salmen, L. F. (1999). *Beneficiary assessment manual for social funds*. Environmentally and Socially Sustainable Development Network.

Salmen, L. F. (2002). *Beneficiary Assessment: An approach described*. Environment Department Papers, Social Assessment Series, Paper No. 010

Jones, B. (1998). *Beneficiary Assessment for monitoring: the Zambia social recovery project*. World Bank, Social Development Notes, Note No. 36

## 2. Most Significant Change

The MSC methodology involves collecting multiple stories from those impacted by the intervention, and then having several levels of stakeholders choose the most significant of the stories and then passing them along to the next level until a small number of the most significant stories have been chosen. The stories are chosen based on how well they highlight the program impact (either positively or negatively). This methodology provides a simple way to understand large amounts of complex data. Through uses of this methodology, programs have been able to identify unexpected changes, provide a rich picture of what is happening, and identify the values of the organization. No professional skills are needed that cannot be learned, so it is a good way to build organizational capacity.

### 1. Pre-data collection activities

The first step of this methodology is to introduce the process to stakeholders and get commitment to participate. Then, domains of change must be established. These are purposefully vague, so that they can be defined by the interviewees. This will include broad domains like “changes in people’s lives.” The last step before data collection is to determine how frequently changes should be probed.

Data collection questions need to be determined. These questions need to be very simple, so that participants can take their answers wherever they feel like it. This is a good way to get a good picture of the intended and unintended outcomes of the interventions.

### 2. Data collection

Stories are collected from beneficiaries. There has been no determination of sample size, but for the initial collection of stories, it should be from a fairly broad range of participants,

usually 100s are collected for the first round. People are asked vague questions such as “In the last month, what was the most significant change that occurred.” These questions are meant to get a broad range of answers from beneficiaries.

Volunteers or staff members can be used to collect data, as long as they have had sufficient training in qualitative interview skills. After interviews are conducted, stories will have to be written up with great detail so that they can be passed along to determine the most significant story.

### 3. Data Analysis

The analysis phase of this methodology is the most intensive. All of the stories that were gathered have to be written up and given to the first set of reviewers. The analysis of these stories is done based on the domains that were established before the interviews began. The first level of the hierarchy of the organization doing the assessment will review the stories that were sent to them and chose the most significant story for each of the domains to send to the next level of the hierarchy, which will then do the same will the stories sent to them. This process will move up the chain until the heads of the organization chose the most significant stories. The figure below demonstrates the process of discovering the most significant stories that was done on a project in Laos.

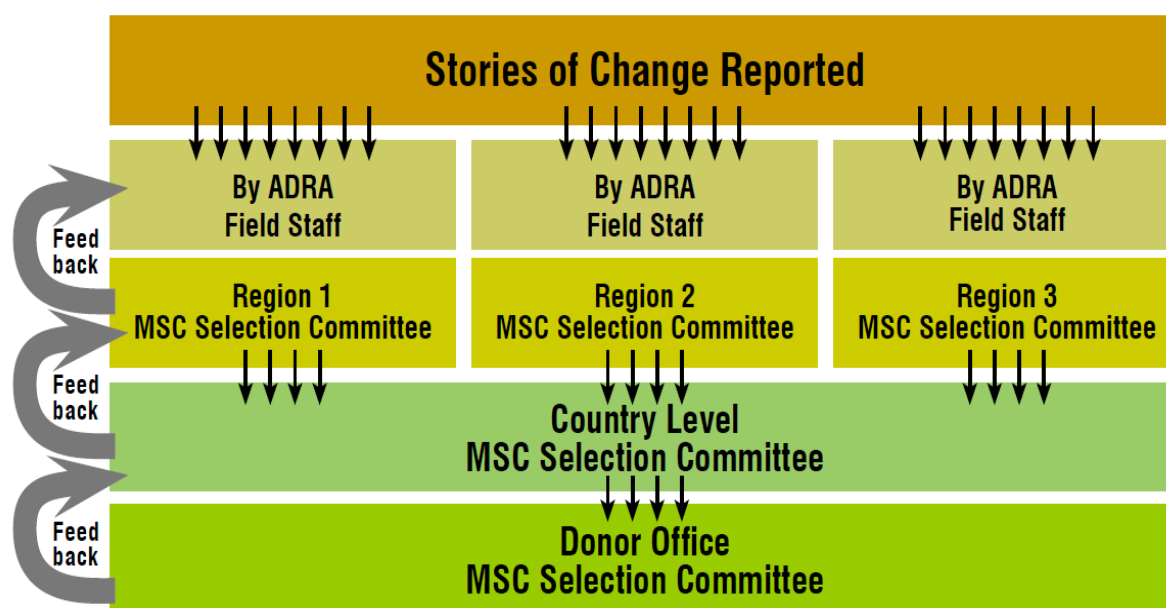


Figure: Most significant change analysis chain (Davies & Dart, 2005)

Each time a story is selected, the criteria that was used and the decision-making process in choosing the story is recorded and then that process is fed back to all levels of the hierarchy. Once the final stories are selected, the stories themselves and the selection criteria used are written up in the final report. The stories are also verified, if possible, to make sure that the information is accurate.

### 4. Evidence Gathered

The evidence gathered using this methodology is mainly beneficiary perceptions of changes in their lives, hopefully in relation to the intervention, and what is of most value to them in these changes. This is better suited for development projects where long term changes over



time are more evident, and may be less effective in rapid onset disasters where everything has changed.

Sources:

Davies, R., & Dart, J. (2005). *The 'Most Significant Change' (MSC) technique: A guide to its use*. Care International, London, UK.

Serratt, O. (2009). *The Most Significant Change technique*. Asian Development Bank

Sango, D. & Dube, S. D. (2014). *Most Significant Change stories: Zimbabwe emergency drought mitigation operation for livestock project*. Smart Development Works.

### **3. Qualitative Impact Assessment Protocol (QUIP)**

QUIP is a methodology that is designed to assess changes that are related to complex projects. Narrative accounts are collected from beneficiaries about changes that occurred and what caused them. It is useful in complex interventions where there is the potential for a lot of factors to influence the outcomes and changes. The defining feature of QUIP is that biases might be encountered if asked directly about the program, so interviewers use open ended, exploratory questions about the changes that have occurred to probe for all changes instead of asking specifically about the program. The interviews should be conducted by someone not associated with the program organization.

#### **1. Pre-data collection activities**

Before data collection, two steps must occur. First, the program staff needs to collect the domains of change and then design the questionnaire around those changes. Domains of change refer to areas in a participant's life that changed and contributed to wellbeing. The questionnaires are created by working backwards from the outcomes of interest, rather than focusing on activities. Impacts that are supposed occur due to the program are based on the theories of change that have been established. It is essential that the questionnaires remain as open-ended as possible so that all changes and everything that contributed to the changes can be recorded.

Different sampling techniques are used based on what the goal of the study is. Typically, selective sampling is used, especially if there are certain over or under performing areas that programs want to explore. Cluster sampling can also be used if different geographical areas have been a part of the intervention and want to be assessed. While there is no concrete sample size used, they are usually planned in sets of 20-25 interviews and four focus groups. This is usually large enough to gather the important information from a cluster or community. No control groups are used in this study.

#### **2. Data Collection**

If at all possible, data collection should be performed by an outside evaluator that has no connection to the program or organization, and limited to no knowledge of the activities of the intervention. Interviews and focus groups will be conducted, usually lasting around 60-90 minutes. All interviews and FGDs have the goal of having respondents mention activities of the intervention without being probed for.

#### **3. Data Analysis**

As opposed to the data collection phase, the analysis needs to be done by someone with strong knowledge of the program and the theories of change. They code for evidence of causal pathways from actions to outcomes taken from both interviews and FGDs. There can be one or more trained analysts working on the analysis. Causal claim statements are highlighted in all interviews and causal maps are created.

#### 4. Evidence Gathered

Evidence is generated on causal claims of the interventions. Outcomes are established and then actions that led to the outcome are discussed. It provides in-depth information on what has changed and how/if the intervention led to those changes. It can gather information on unintended changes/reasons for changes.

Sources:

Copestake, J. (2014). Credible impact evaluations in complex contexts: confirmatory and exploratory approaches. *Evaluation*, 20(4), 412-427. DOI: 10.1177/1356389014550559

Copestake, J., Allan, C., van Bekkum, W., Belay, M., Goshu, T., Mvula, P., Remnant, F., Thomas, E., & Zerahun, Z. (2018). Managing relationships in qualitative impact evaluation of international development: QUIP choreography as a case study. *Evaluation*, 24(2), 169-184. <https://doi.org/10.1177/1356389018763243>

#### 4. General Elimination Method

This methodology is a way to determine causality in impact assessments without doing an experimental study, by identifying all possible explanations for change and then systematically eliminating each explanation until only the true explanation is left. This is done by interviewing both beneficiaries and stakeholders to determine what all possible explanations for changes are and collecting large amounts of background information. Local conditions and historical events are probed for other explanations for change.

##### 1. Pre-data collection activities

Before interviews are conducted, it is essential that as much information is collected on the outcome and intervention as possible. This typically includes all of the monitoring reports available, any documents on intervention start up, and any information on historical events and local conditions. This data will need to be synthesized and analyzed to determine what other information is needed to make judgements on eliminations.

##### 2. Data Collection

Data collection for this methodology takes place in the form of interviews. These interviews will be with both intervention participants as well as other stakeholders in the intervention and people that have a good understanding of the community. The number of interviews will vary depending on what information is needed. These interviews will be conducted by either volunteers or staff members that have been trained in qualitative data collection.

##### 3. Data Analysis

Data analysis will include qualitative analysis of the interviews as well as the synthesis of the qualitative data collected with the other information collected. Once all of the information has been collated, the analyst will need to look at all scenarios that might have led to the outcome, and systematically eliminate all possibilities that can be eliminated using the

evidence gathered. The goal is to have the intervention be the possibility left after all others have been eliminated.

#### 4. Evidence Gathered

While this method provides a lot of good information on whether the intervention actually resulted in the impact observed and helps determine causality, it gives limited information on individual beneficiaries.

Sources:

White, H. & Phillips, D. (2012). *Addressing Attribution of Cause and Effect in Small-n Impact Evaluations: Towards an integrated framework*. Working Paper 15, 3ie.

Leeuw, F. & Vaessen, J. (2009). *Impact evaluation and development: NONIE guidance on impact evaluation*. Network of Networks on Impact Evaluation.

#### 5. Outcome Harvesting

Outcome harvesting is used in retrospective evaluations of complex systems to help make sense of out of outcomes after they emerge. Either staff or evaluators identify, formulate, verify, analyze, and interpret outcomes in contexts where cause and effect are not well understood. Who changed what, why it matters, and how the program contributed to the outcomes are assessed. This is usually done using FGDs and interviews. This methodology looks at what has been achieved, and then works backwards to determine whether the interventions played a role in the outcome.

##### 1. Pre-data collection activities

The first step in outcome harvesting is to develop agreements with the people that will use the results and decide what the priority question for the harvesting will be. Users and researchers also have to agree on the process, what questions will be asked, from whom, etc.

Outcome descriptions are then created, the level of detail depending on the organization and the harvesting questions. They can range from a sentence to a full-page description. They include information such as the context of the outcome, perspectives on the outcome, and importance. To create these outcomes, all available data and documentation from the program is examined.

##### 2. Data collection

Once the outcome descriptions are generated, interviews need to take place to review the outcome descriptions and potentially formulate new outcomes that were not thought of. These interviews will be with both people that experienced the intervention, as well as people that worked on the intervention.

Once this information is obtained, all of the outcomes that have been determined are looked at again and the most appropriate are picked to be verified. The verification process allows an increase in accuracy and credibility of findings. This verification is done by giving the information to independent observers who are knowledgeable about the outcome and the contribution of the change.

##### 3. Data analysis

Once the final outcomes have been verified, they are then classified into similar outcomes. The causes that led to the outcomes are also analyzed. A database may need to be created to organize the outcomes if a large number of outcomes are collected. Stories, charts, and visualizations are different ways that the outcomes can be analyzed. The goal is to see how different outcomes contributed to a process of change.

#### 4. Evidence

This methodology provides evidence on what outcomes are achieved, and how the intervention contributed to these changes, and what the outcome means for the intervention. This provides a better representation of overall program effectiveness than individual beneficiary impact assessments.

Sources:

Canto-Blundo, G., Laderach, P., Waldock, J., & Camacho, K. (2017). Learning through monitoring, evaluation and adaptations of the "Outcome Harvesting" tool. *Cahiers Agricultures*, 26.

Holvand, I. (2005). *Successful communication: A toolkit for researchers and civil society organizations*. Overseas Development Institute Working Paper 227. London

Wilson-Grau, R. & Britt, H. (2012). *Outcome Harvesting*. Ford Foundation's Middle East and North Africa Office.

### Mixed-Methods:

#### 1. MAPP (Method of Impact Assessment in Projects and Programs)

MAPP is a heavily participatory methodology that is designed to evaluate the impact of interventions and allow beneficiaries to express ideas on how interventions can be improved. It was originally designed to assess development programs in the agriculture sector, but has now been used in multiple assessments. This methodology provides an open approach that can evaluate both planned and unplanned impacts. A specific program is evaluated in relation to other ongoing programs, and beneficiaries have strong inputs into all aspects of the assessment. It is mainly qualitative, but some of the qualitative data is turned into quantitative information such as point systems for ranking changes and impacts. Using the data gathered, and influence matrix is created to try to determine attribution.

##### 1. Pre-data collection activities

The first step of this methodology is to determine who will be leading the workshops. This person or team will need to undergo a training on the methodology. This training lasts a day. Once the evaluators are trained, group discussions are organized. These discussions usually involve both beneficiaries and non-beneficiaries, with different sexes and socio-economic classes included. These people need to be identified. There is no set size for these discussion groups.

##### 2. Data collection

Data is first collected from beneficiaries and non-beneficiaries on the effects of the project using developed tools. Project measures and activities are identified, and it also looks at contributions made by individual development measures on the overall developments. There are six different tools that have been established, and each of the discussions have beneficiaries fill out all of the tools in a pre-defined order. These tools include things such as

lifelines, trend analysis, list of measures, and influence matrix. All of this information is summarized into a development and impact profile. This process can happen over a two-day period.

### 3. Data analysis

Analysis is performed by both participants in the discussion and the discussion leaders. Participants fill in the development and impact profile using what information was gleaned from the trend analysis and influence matrix, where positive and negative trends are determined. This will be monitored by the discussion leader, but run entirely by the participants.

The other qualitative data that is collected during the discussion will be analyzed by the discussion leaders to provide more context to the data.

### 4. Evidence

MAPP gathers evidence on the changes that have occurred over time and the reason for these changes. It also gathers information on which parts of programs had the biggest impact.

Sources:

Neubert, S. (2010). *MAPP: a participatory method for impact assessment of programs and projects*. German Development Institute.

White, H. & Phillips, D. (2012). *Addressing Attribution of Cause and Effect in Small-n Impact Evaluations: Towards an integrated framework*. Working Paper 15, 3ie.

## 2. Participatory Impact Assessment

This approach attempts to answer the following questions: what changes have there been since the start of the project and are the changes attributable to the project? To answer these questions, both interviews and FGDs are conducted. This uses control groups, develops baseline data, triangulates data, and uses standardizing and repeated methods to answer these questions. Data is collected in qualitative techniques, but some data is quantified to be used for quantitative analysis. There are eight stages to this methodology: define the questions to be answered, define the geographical and time limits on the project, identify locally prioritized impact indicators, decide which method to use for measuring change, decide sampling method and sample size, decide how to assess attribution, decide best way to triangulate data, and plan feedback.

### 1. Pre-data collection activities

Prior to any other activities, it must be decided what questions to answer in the impact assessment. There should not be more than five questions to answer from the assessment, and fewer questions will result in more detailed data. Step 2, establishing timing and geography, will make sure that everyone understands the limited of the interventions and more importantly the scope of the impact assessment.

For the impact assessment to take place, indicators of impact need to be established. Impact indicators measure the changes that have occurred due to the project. This is typically some sort of change in livelihood. Additionally, there are several different methods used to

measure these indicators. These include things like ranking and scoring, before and after scoring, proportional piling, and matrix scoring among others. All methods using interviews to get the data, but some of the interview techniques produce quantitative.

Sample sizes for interviews and FGDs need to be established. This has been done differently in different assessments using this methodology. Random, purposive, and convenience sampling have all been used. The type of sampling is determined by multiple different factors, including the amount of money that has been allocated for the assessment. Random sampling produces the largest sample size, so that is only done if time and money permits.

## 2. Data collection

Data is collected from both interviews and FGDs. The format differs depending on the program being evaluated. Commonly, beneficiaries and non-beneficiaries are both interviewed or there are focus groups with both to compare those that did and did not receive the intervention. Most of the data collection instruments have already been designed, they just need to be modified for the individual intervention. Both volunteers and staff members can collect data, as long as they have been trained in the methodology and qualitative data collection techniques.

## 3. Data analysis

Although mostly qualitative techniques are used to collect data, the analysis involves both qualitative and quantitative analysis techniques. The goal of collecting the data is to try and prove attribution, and therefore data is analyzed with that goal in mind. The data also needs to be triangulated between different methods and secondary data can also be collected and analyzed to assist in triangulation.

## 4. Evidence gathered

Evidence on differences in the indicators between beneficiaries and non-beneficiaries is gathered, along with baseline data (if not previously collected), evidence of attribution, and case studies to determine individual impact. Perceptions of impact of the program on individuals is generated. Additionally, numeric evidence on changes in indicators is gathered. Once all data is gathered, it is shown to the community. This is an opportunity to verify that the results of the impact assessment are accurate and to add even more context to the data if necessary.

Sources:

Catley, A., Burns, J., Abebe, D., & Suji, O. (2013). *Participatory Impact Assessment: A Design Guide*. Feinstein International Center, Tufts University, Somerville

## 3. Citizen Report Card

The ideas behind the Citizen Report Card is that individuals should have a say in the way that services are provided. They can provide good information on impact of services, as well as adequacy of the overall program and problems in service delivery. This methodology is mainly used in evaluation public services, but can also be used to evaluate NGO

interventions. Data from the report card is meant to reflect the lived experiences of people, as well as the impact of these services.

#### 1. Pre-data collection activities

In order to conduct these surveys, focus group discussions need to be designed to determine the purpose and scope of the study. These interviews can be with government official, NGOs, citizen groups, and agency heads. Once the scope of the study is determined and questions to be assessed are created, a questionnaire needs to be developed to give to participants. Sampling and sample size determinations will need to be done, but the sample size should be large for this technique.

#### 2. Collection

Both qualitative and quantitative information is collected using this methodology, so it will be important to have somebody with knowledge of both techniques. The FGDs will collect data on the experiences of officials to help guide questionnaire and study design. Someone will need to facilitate this data collection. Quantitative surveys will then need to be sent out to participants. Large amounts of data should be collected on how participants interacted with the services provided.

#### 3. Analysis

Both qualitative and quantitative analysis will need to be done, so it is important to have someone available who understands and can do both types of analysis. This is a general community survey, so it could be possible to do a comparison between beneficiaries and non-beneficiaries depending on design and sampling.

#### 4. Evidence gathered

This is more a tool to measure the quality of the services that were provided and the satisfaction with these services. These questions might be able to be tailored to the project, and would provide some good information on whether beneficiaries were satisfied with what they received, but it would not measure impact in the way it is currently designed.

Sources:

The World Bank. (2004). *Citizen report card surveys: A note on the concept and methodology (Note No. 91)*. World Bank Social Development Notes.

Ravindra, A. (2004). *An assessment of impact of Bangalore citizen report card on the performance of public agencies (ECD working paper series 12)*. The World Bank Operations Evaluation Department

### 4. Collaborative Outcome Reporting

This is an impact assessment methodology that is based around a performance story that presents evidence on how a program has contributed to outcomes and impacts. This technique maps out existing data and additional data against the program logic to produce this story. There are six steps used in this approach: 1. Scoping inception and planning workshop, 2. Data trawling, which looks at previously collected data, 3. Social inquiry, which involves collecting new data, 4. Data analysis and integration, 5. Outcome panel, and 6. Summit workshop. Steps 5 and 6 involve expert panels that check the credibility of the evidence surround impacts and the extent to which they can be attributed to an intervention.



## 1. Pre-data collection activities

The pre-collection phase involves a scoping/inception workshop. This is held so that stakeholders can have a say in the planning of the intervention. The program logic is developed during these workshops. This is what is used to determine how the intervention should theoretically work, and how the intervention will lead to the impact. Sources for existing data are also identified. These sources could include existing census data, routine monitoring data, program baseline data if collected, along with other sources. Evaluation questions that will be asked during the social inquiry are also developed.

Sampling for this methodology will depend on the type of data that is collected. Typically, qualitative information is gathered on stories of change (using techniques similar to MSC). Additional data can be collected, including quantitative survey data. Different sample sizes will be needed for different techniques.

## 2. Data collection

There are two stages of data collection, the data trawl and the social inquiry. The data trawl includes both primary and secondary data. Program staff can be used to help discover and collate the data. The goal of the data trawl is to look for existing evidence on the program and its contributions to the outcome. The data trawl can happen in confluence with the social inquiry. The social inquiry will involve collecting data from the participants in the program. This can be either quantitative or qualitative, but usually it is a combination of both.

## 3. Data Analysis

Both quantitative and qualitative analysis will need to be done. This will require staff members or outside consultants with knowledge of both techniques. Results charts are usually used to collate the different types of data.

After the data is collated, people with relevant knowledge will be brought together to assess the results and determine the contribution that the intervention had on the outcomes. They also look at rival hypotheses that might also explain the outcomes. A final workshop with staff members, community members and other stakeholders key findings and recommendations are presented and synthesized. Examples of change are identified and added to the evidence base.

## 4. Evidence Generated

This methodology creates a performance story that presents evidence of how the program contributed to outcomes of interest, with input from both the community and technical experts. It collects perceptions of how the program contributed to wellbeing/outcomes during the disaster. There has been criticism that this technique produces overly positive reports.

Sources:

Dart, J. & Roberts, M. (2014). *Collaborative outcomes reporting*. BetterEvaluation. Retrieved from <http://betterevaluation.org/plan/approach/cort>

Roughley, A., & Dart, J. (2009). *Developing a performance story report: User guide*. Commonwealth of Australia

## 5. Success Case Method



**This technique is described in depth in Appendix F and the main body of the report.**

## **6. Case Study**

A case study focuses on a particular unit-person, site, project to focus on how different elements fit together and how different elements have produced the observed impacts. There are several different types of case study methods, but this will look at the program effect method. Program effects methods examines causality and usually involves multimethod assessments. Case studies entail analysis of a complex situation, based on a comprehensive understanding of the instance obtained through extensive descriptions and analysis of the instance. It is most typical to look at the project as a whole or more than one project for analysis.

### **1. Pre-data collection activities**

The first step in this methodology is to determine which unit will be used as a case. This could range from individual, household, community, region, or program. In most FbF interventions, the cases should be individuals or communities. All the data that is available should be collected, including any existing data. This needs to be done before data collection, as data collection should be focused on what is missing from existing data. This data might come from any baseline info gathered, monitoring reports, census data, and any other data sources available. This can be a lengthy process, as it is necessary to gather and synthesize all available data.

### **2. Collection**

Once the holes in the existing data have been identified, it is necessary to collect data to fill in those gaps. This is usually done by conducting in-depth semi-structured interviews with the cases (if cases are individuals). Staff members or volunteers can collect data. Different data will need to be collected on different cases, depending on what existing data was available for each individual.

### **3. Analysis**

Both collected and existing data will be analyzed, so it will be essential that someone is available that has both qualitative and quantitative analysis skills. Both types of data will be fairly straightforward to analyze. The more difficult analysis is the synthesis of all data. Data needs to be triangulated and synthesized to present a coherent case.

### **4. Evidence Generated**

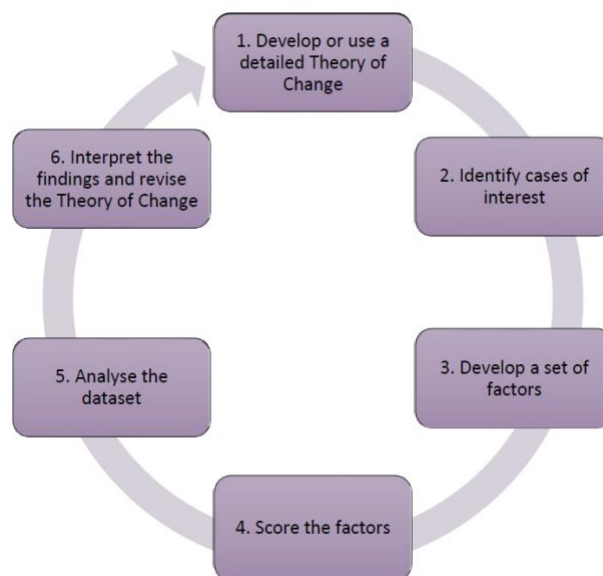
Evidence will show the effects of the program and reasons for success and failure. This method gives good information on the program overall, and if it was successful, which is good information to have for the Red Cross to evaluate the program as a whole, but does not provide reliable measures of the benefits for a household. Depending on what unit is used, the focus may not be on individual households at all or may not be generalizable to households.

Source:

Balbach, E. D. (1999). *Using case studies to do program evaluation*. California Department of Health Services.

## 7. Qualitative Comparison Analysis

QCA involves analysis of multiple cases in complex situations to look for patterns and help explain why change happened in some situations and not others; systematically generating findings across multiple case studies using qualitative and quantitative data. It is based on two assumptions: change is often the result of different combinations of factors, and different combos of factors can create similar changes. Meant to be a rigorous process that follows the following structure: 1. develop theory of change, 2. identify cases of interest, 3. develop factors, 4. score factors, 5. analyze data, 6. interpret findings and revise theory of change. Below is a diagram that shows the structure of the QCA.



*Figure: QCA methodology steps (Simister & Scholtz, 2015)*

This methodology is case-based, and involves generating findings across multiple case studies. Both quantitative and qualitative data is gathered from between 10 to 50 cases. The basics of the methodology is to analyze patterns across multiple case studies to better understand why some changes happened and others did not.

### 1. Pre-data collection activities

The first step of the methodology is to develop a theory of change (or use a pre-existing one). The theory of change identifies two things, the change the study is interested in and the factors that can bring about that change. This change to be analyzed is usually the outcome of interest. The next step is to identify cases. Some of the cases should have the outcome of interest and others should not, but they should be similar in other respects. Cases can be anything from different governments to individuals, but for FbF impact assessments they will always be individual beneficiaries. Factors also need to be established before data collection. Factors are aspects whose presence or absence contribute to the outcomes.

### 2. Collection

Once the factors are identified, intensive qualitative data needs to be collected on the cases. Any available data should be collected, and new data should be collected as well. For FbF

interventions, this would require in-depth qualitative interviews with each of the cases. Collected data should focus on factors and outcomes.

### 3. Analysis

Once the data is collected on the cases and factors, the factors need to be scored. This involves looking at the factors across all of the cases and determining a score. Typically, this means looking at a factor across all cases and scoring it either a 1 if the factor is present or 0 if it is absent. Once the scoring has occurred, data needs to be analyzed. This can be done by eye if relatively few cases, but it is usually done using software. Analysis usually looks at what combination of factors or absence of factors led to the outcome. After the data is analyzed, it is essential to go back to the cases to see if the findings make sense.

### 4. Evidence generated

This methodology provides numerical and qualitative evidence on different pathways that led to the outcome of interest. It gives data on factors that are associated with positive outcomes of interest. It provides very detailed household level data on impacts.

Sources:

Schatz, F. & Welle, K. (2016). *Qualitative Comparative Analysis: A valuable approach to add to the evaluator's toolbox? Lessons from recent applications*. CDI Practice Paper No. 13.

Simister, N. & Scholz, V. (2017). *Qualitative Comparative Analysis*. INTRAC.

White, H. & Phillips, D. (2012). *Addressing Attribution of Cause and Effect in Small-n Impact Evaluations: Towards an integrated framework*. Working Paper 15, 3ie.

## 8. Realist Evaluations

The overall goal of this method is to identify the underlying generative mechanisms that explain how the outcomes were caused and the influence of context. The reasoning of the beneficiaries in response to resources or opportunities provided by the intervention is what causes the outcomes. First, the program theory is developed, then data is collected in whatever manner suits the question, then the data is organized in relation to the initial theory, then patterns of outcomes are analyzed, then the mechanisms generating those outcomes can be analyzed. Finally context-mechanism-outcome statements are created. It is particularly good at determining underlying causation and why different outcomes are achieved in different contexts.

### 1. Pre-data collection activities

The first step of realist evaluations starts with developing the program theory. The program theory describes how the intervention is expected to lead to the outcomes and in what conditions these outcomes occur. The difference between realist and other types of evaluations is that a realist theory details the mechanisms that will generate the outcomes and what features of context will determine whether or not these mechanisms will operate. The three elements, mechanism, outcome, and context are specified in the pre-collection phase. Data collection is then focused on testing the different elements of the program theory.

### 2. Collection

Typically, both qualitative and quantitative data is collected, with quantitative data focused on context and outcomes, while qualitative data to look at generative mechanisms. This collection typically uses the case study design, because cases can test the mechanism, outcome, context statements. Usually purposive sampling is used, and no sample size is specified.

### 3. Analysis

Analysis depends on the type of data that is collected. For the overall methodology the analysis is done under the context-mechanism-outcome (CMO) frameworks. In the first phase of the analysis, the data is organized around the program theory. This involves either qualitative coding, or basic quantitative analysis depending on what was gathered all centered around the CMO framework. Outcomes data also needs to be disaggregated based on sub-groups. If outcome patterns are identified, mechanisms generating the outcomes can then be analyzed. Once the mechanisms are analyzed, contexts in which these mechanisms did and did not work can then be determined. Contexts could be either different sub-groups, processes of implementation, socio-economic conditions, or other factors. Lastly, analysis needs to be done to determine which CMO pathway offers the most plausible explanation for the observed outcomes.

### 4. Evidence gathered

Evidence is generated on how and why the program works, and for which people it works best. It should give details on the contexts in which a program works best. This method will give information on who the program is working best for, in which context, and the mechanisms of how the program works. This is all good data to have, but it is less focused on individual effects and more focused on the overall program theory.

Sources:

Westthorp, G. (2014). *Realist impact evaluation: An introduction*. Overseas Development Institute.

White, H. & Phillips, D. (2012). *Addressing Attribution of Cause and Effect in Small-n Impact Evaluations: Towards an integrated framework*. Working Paper 15, 3ie.

## 9. Process Tracing

The main purpose of process tracing is to attempt to establish whether, and how, a potential cause or causes influenced a specific change or set of changes. This is done by applying a set of formal tests to examine the strength of evidence. A key feature is development and testing of alternative ideas of how change occurred. process tracing can show whether a change occurs and how and why it occurred as well. It works by first identifying the change of interest, identifying the evidence that confirms that change has happened, documenting the processes that have led to change, establish alternative causal explanations, and then assess all hypotheses. These evaluations are done using the formal tests provided in academic literature.

### 1. Pre-data collection activities

The first step is to identify the change that occurred that are of interest. It is difficult to do process tracing with multiple changes, so stakeholders will need to work together to determine which change that occurred is the most important to analyze. After the change

has been established, evidence that the change occurred needs to be gathered. This can be gathered from documents and collecting data. The data collected can be qualitative or quantitative. Once it has been established that change has occurred, processes that led to change need and alternative causal explanations need to be developed.

## 2. Collection

Data needs to be collected at two times in this methodology. First, data needs to be gathered to establish that change has occurred. Secondly, evidence needs to be collected on the hypotheses that were established about the causal explanations. There is no explicit directions for what type of data needs to be gathered, it just needs to establish a change and evidence that can strengthen or weaken hypotheses. Depending on what type of data is collected, enumerators will need to be trained in either qualitative or quantitative data collection.

## 3. Analysis

Analysis will need to be done three times for this methodology. First either qualitative or quantitative data will need to be analyzed to establish that changes occurred, then analysis will need to be done on the data that establishes causal explanations, then the formal tests used to evaluate the evidence will need to be used. Depending on the type of data that is collected, someone with expertise in either qualitative or quantitative evaluation (or both) will be required.

## 4. Evidence gathered

At the end of the process, an evaluator should be able to assess the extent to which competing causes or hypotheses may or may not have contributed to a change or set of changes. They should also be able to state how and why the change or changes came about. This method gives good information on the causality of the program to the impact, as well as plausible explanations for why things changed, but focuses more on the overall program than on the individual beneficiaries

Sources:

Anguko, A. (2019). Process Tracing as a methodology for evaluating small sample size interventions. *eVALUation Matters*, Second Quarter 2019, 19-27.

Punton, M. & Welle, K. (2015). *Straws-in-the-wind, Hoops and Smoking Guns: What can process tracing offer to impact evaluation?* IDS CDI Practice Paper, No. 10, 2015.

## 10. Contribution Analysis

Contribution Analysis explores the contribution a program is making to observed results by gathering evidence and creating a contribution story. This approach was developed to explore the attribution of the program to the impact where experimental or quasi-experimental programs are not possible. IT relies heavily on analyzing a programs theory of change and looking at any other potential contributing factors. There are 6 steps to the method: Set out attribution problem, develop theory of change and risks to it, gather existing evidence, assemble and assess the contribution story, gather new evidence, and revise and strengthen contribution story. Contribution Analysis attempts to explore plausible association, to ask “whether a reasonable person, knowing what has occurred in the

program and that intended outcomes actually occurred, agrees that the program contributed to those outcomes.”

### 1. Pre-data collection activities

Most of the steps of this methodology occur prior to data collection. These include establishing the attribution problem, developing the theory of change and risks to that theory, gather existing evidence, and assembling and assessing the contribution story. Establishing the attribution problem involves understanding what has occurred during the intervention and acknowledging that other factors besides the intervention could have caused or contributed to the outcome. Developing the theory of change and risk to the theory involves creating a logic model for the intervention, having a good understanding of social and economic issues that might affect the outcome, having realistic expectations of how the program can contribute to the outcome, and considering links between activities. Gathering existing evidence involves finding all of the evidence that has been generated throughout the intervention, identifying key stakeholders, and identifying gaps in the evidence. The last step pre data collection is to start creating a contribution story, and assessing the strengths and weaknesses of the intervention steps in the program logic. This step also looks for other influencing factors besides the intervention that might contribute to the story.

### 2. Collection

Data collection for this methodology is used to add data to the contribution story. This is most likely to be qualitative data, but may also include some quantitative data. This data will be used to confirm or refute the information that has already been established in terms of theory of change and causal pathways.

### 3. Analysis

This data gathered is then used to strengthen and revise the contribution story. The type of analysis will depend on whether qualitative or quantitative data is gathered.

### 4. Evidence generated

The final picture established from this analysis should be able to tell whether expected results were due to the theory of change, and that the program was influential in bringing about those results. This technique gives very good information on whether an observed result is attributable to a program. There is less focus on individual beneficiaries and more emphasis on the program. It does give information on whether the program accomplished it's theory of change, and of that theory of change is about helping individual beneficiaries then that will give you info on household level effects.

Sources:

Mayne, J. (2001). Addressing attribution through Contribution Analysis: Using performance measures sensibly (discussion paper). Auditor General of Canada.

Mayne, J. (2008). Contribution analysis: An approach to exploring cause and effect (Brief 16). The Institutional Learning and Change Initiative.

## Appendix D: Additional Interview Quotes

Current Practices	Challenges
<p><b>Data Collection:</b>  <i>“OK, first of all we collected the socio-economic status before and after the event, and their present condition. We also collected a few information regarding socio-economic condition like number of livestock, assets, debt that were before the disaster and their situation right now. That is before and after the situation. And we collected information for the beneficiary group as well as the control group. So we understood what the impact was for our group.”</i></p> <p><i>“First, we have to plan for the data collection methods. Whether we will go to the fields or collect over the phone.”</i></p> <p><i>“And it took about 2-3 days in each district to collect the data. And we selected 3 districts.”</i></p> <p><b>Control Groups:</b>  <i>“For floods we selected control groups based on our existing data, and we didn’t provide them care. They had a vulnerability score for the family in preparing the cash intervention. So we selected for flood, the number of beneficiaries that were based on the vulnerability score of the family. Based on one or two standard deviations, where it was not possible to find one standard deviation away.”</i></p> <p><i>“We collected control groups data from those who went to the shelter and didn’t provide support at the shelter. At the same time, we selected some control from the population that didn’t go to the shelter.”</i></p> <p><i>“Yes, so that was a random selection. We know where our beneficiaries are from, we have their details, but the non-beneficiaries they are living in nearby communities or in the same communities but from different pockets. And they were randomly selected.”</i></p> <p><b>Technology:</b>  <i>“And what we’ve seen, in terms of efficiency, using the mobile phones is best. Kobo or Kobo collect. WE are now using this, but in the past the red cross has used so many tools that have been developed on hard paper.”</i></p>	<p><b>Disjointed Study Design:</b>  <i>“But that disconnection on the way that its done, I think its complicated. So you have the first group that does the data collection, and then you have someone else takes care of the data analysis, and then you have someone else interpreting the results.”</i></p> <p><i>“The fact that they are seen as independent process, then you cannot go back and explore more into why or analyze the data as it comes in.”</i></p> <p><b>Lack of Capacity:</b>  <i>“this is very cost intensive and difficult and I don’t know how we are going to pull this off.”</i></p> <p><b>Lack of Qualitative Data:</b>  <i>“But it would better to get information from a FGD, that would describe more rational.”</i></p> <p><i>“methodology also covers more subjective topics and not just the pure hard numbers but also what else was there that we could rescue from this intervention.”</i></p> <p><i>“And I think one of the challenges of the impact assessments that are being done is that it does not take into account the kind of broader contextual issues that exist and the national society starting points as well.”</i></p> <p><b>Difficulties with Sampling:</b>  <i>“The scenario is different during flood and dry period. We can directly drive to their houses during flood, but during dry period we have to walk up to 10 km to find them. That was very challenging.”</i></p> <p><i>“this year, we selected a few households for sampling, but after going to the field we found that they are not in their houses.”</i></p> <p><b>Timing:</b>  <i>“The other is you know, the whole thing about the timeliness. This is the main problem that we have faced. The whole thing about who is this for and how long does it take”</i></p>

Capacity	Information to Collect
<p><b>General Comments:</b>  <i>"I think there should always be a capacity building element to it, so that then Mongolia red cross can feel like, maybe not the next time, but that they are able to really drive this"</i></p> <p><i>"their partners are often international consulting groups. Which it's fine to work with those partners, but it would be good to build more local capacity on that."</i></p> <p><b>Community and Volunteers:</b>  <i>"We have a network of volunteers at the red cross, and some of the volunteers have been trained in the FbF methodology."</i></p> <p><i>"we collected 440-480, and for flood we collected 446 number of information. And we had almost 100 volunteers."</i></p> <p><i>"they're also our networks at the community level and our entry into the community. So we leverage on our volunteer network on everything. Not just the after action reviews, but all of our programming"</i></p> <p><b>M&amp;E, Staff, and Statistics:</b>  <i>"We need some bit of capacity building. Such that even none M&amp;E staff can do any of these big surveys because it might be hard for the NS to recruit some M&amp;E people"</i></p> <p><i>"Yea, they do the data analysis part also. Because we invest in them and they are doing everything: that is data collection framework, sampling design, field testing as well as data collection. And those data processing is done by them. And they will analyse and evaluate the heat wave simulation data"</i></p> <p><i>"However, internally I can say that we have an above average capacity in terms of statistical analysis, and human resources to do these types of analysis."</i></p> <p><i>"For this assessment, one person from the national headquarters, he went down to the district and oriented the volunteers and there we included at least 25 to 30 volunteers and they were engaged"</i></p>	<p><b>Qualitative Data:</b>  <i>"I am not sure they're really is a replacement for some good qualitative interviews."</i></p> <p><i>"For these qualitative interviews...even though it does not give us any statistics, it gives us an idea of the types of benefits and who was receiving them and why."</i></p> <p><i>"We also keep their satisfaction from before and after, which gets more qualitative information. How they are satisfied."</i></p> <p><i>"It should be qualitative, otherwise it will be valueless."</i></p> <p><b>Outcomes of Interest:</b>  <i>"we should put the emphasis on the type of activities that we are doing and the long term impact that they are having"</i></p> <p><i>"Not being too heavy, but we want to know that the assistance is getting to the right place at the right time, and if it isn't we need to know what we can do better."</i></p> <p><i>"I would like to see how our support benefits them to save their livelihoods and lives."</i></p> <p><i>"But overall we just want to see that there is still access to services, we don't lose lives or livelihoods, there are no disease outbreaks, and people carry on as usually"</i></p> <p><i>"When we ask national societies what they would like to know, we would also be able to capture those unexpected outcomes."</i></p> <p><i>"start to compare like with like, rather than just people who receive assistance to people that don't receive assistance."</i></p> <p><b>How the program is working:</b>  <i>"we should put the emphasis on the type of activities that we are doing and the long term impact that they are having"</i></p>



Information to Collect (Cont)	Novel Data Collection Ideas
<p><b>How the program is working:</b>  <i>"One of the aspects that we want to see covered are the institutional aspects. You know managing the case for evaluation in Mongolia and Bangladesh we'd like to slip some of those questions in to assess that as well."</i></p> <p><i>"So how many people are displaced, how many are affected, and then how many households have been damaged, and other sources of livelihoods such as gardens, crops, and then infrastructure to see if the roads are accessible if we are to do any responses in the future."</i></p> <p><i>"So looking for justification is really our first priority and my main expectation"</i></p> <p><i>"Making sure that what we are saying when we develop these early actions is indeed true, that we are making a lot of theoretical explanations. Some are based on already a few evidences that we have collected in small scale activations and tests, but there is still a lot that we need to verify."</i></p> <p><b>Other Data:</b>  <i>"I mean I think there are an infinite number of questions for what action, for what hazard, in what location, with which lead time does that work"</i></p> <p><i>"And I know that to do that you also need to do a rapid assessment on how the markets are functioning"</i></p> <p><i>"I am fighting for the idea of having something that we can all evaluate and report to the government together."</i></p> <p><i>"what is safe to assume? If you know enough, is it safe to assume that if you give people cash, in these types of contexts with these types of lead times that is will have this kind of benefit."</i></p> <p><i>"If we manage to have a standardized lessons learned template for early action interventions and we said look, this is what we could have and this is what you should do and you should produce a report at the end."</i></p>	<p><b>Cellphones:</b>  <i>"there have been some studies on mobility using cell phone, aggregated cell phone information. So you can see if everyone on the coast of Bangladesh starts to go to Dhaka to get jobs after a cyclone, you know and they come back only on the holidays to see their family on the coast, that's something you can measure using aggregated cell phone data."</i></p> <p><i>"There are some techniques out there to randomly poll people via phone call or text message."</i></p> <p><i>"In Uganda we were interested in collaborating with UReport, which was a polling thing set up by UNICEF that would you know, we could send a text message to everyone signed up in a geographical region, so we were going to have our volunteers like help sign people up to this thing."</i></p> <p><b>Remote Sensing:</b>  <i>"So there's a data set from NASA that records night light. Which is often used as a proxy for GDP, or proxy for economic activity, so if you could say ok well there's this drought that is predicted across the region and we were able to reach half the people in that region. And then you monitor the night lights over time and you could show that the region that received the intervention was brighter and showed more economic activity than the region that didn't, you wouldn't actually have to do a lot of work to do that."</i></p> <p><i>"He was proposing satellite images to look in areas that we do house strengthening or early harvesting of crops and it would compare with the neighboring community to see perhaps a different type of impact."</i></p> <p><i>"And in that case, the idea of having satellite imagery, or something that does not require us to talk to beneficiaries may be able to help with objectivity."</i></p>

## Appendix E: Scoring Rationale

- RCT:
  - Capacity required: need very strong statistical skills for both the sampling and the analysis. Also need good M&E experience for questionnaire development, sampling, etc. Large number of volunteers to train
  - Preference: All programs seemed capable of doing quantitative data gathering and said it was common practice, but it does not fulfill what they want in terms of qualitative data gathering. They did have issues reaching some of the sampled people in the past, so such a rigid design could be difficult
  - Complex: very complex sampling and study design, especially in emergency settings. Ethically complex as well
  - Cost: Some interviewees mentioned it was costly to have so many volunteers in the field for a long time collecting data. Also without the capacity to do these big studies, several programs hired consultants to help, which can be costly
- Quasi-experimental:
  - Capacity: need very strong statistical skills for the matching, sampling techniques. Also need good M&E experience for questionnaire development, sampling, etc. Large number of volunteers to train
  - Preference: All programs seemed capable of doing quantitative data gathering and said it was common practice, but it does not fulfill what they want in terms of qualitative data gathering. They did have issues reaching some of the sampled people in the past, so such a rigid design could be difficult
  - Complex: Still complex, with maybe a little less complexity than the RCT because it is a bit less rigid in design. Depending on how non-beneficiaries are chosen it can also be less ethically complicated
  - Cost: Most of the same costs incurred as the RCT design. Potential for a little less money needed on volunteers if sampling is a bit more relaxed.
- Contribution Analysis:
  - Capacity: Work needs to be done for this before data collection, including cause and effect question generation, prior evidence gathering and analysis. This could require moderate capacity from M&E team. Also either questionnaire or semi-structured interview questions need to be developed
  - Preference: There are two different types of data collection, either is fine with NS, although they prefer both, which is why this was given a 3
  - Complexity: The actual data collection phase is not very complicated, there are no comparison groups. The complication comes from the pre-data gathering stage. Depending on the type of data gathered, the analysis stage is not prohibitively complex either
  - Cost: This depends on what type data is collected, but the cost will be less than a large RCT or quasi-experimental study because the sample size will be smaller. Unless there is very very limited capacity, it doesn't seem like this would require hiring a consultant.

- Realist Evaluation:
  - o Capacity: Someone from the national society (or a consultant) will need a very strong understanding of the methodology, theory of change, will need to develop content-mechanism-outcome statements, and may need to run workshops with stakeholders before data collection. Questionnaire development and sampling will also have to take place. This will require moderate to high capacity of the NS team. Data analysis will also need to be done
  - o Preference: This can involve mixed-methods, both qualitative and quantitative collection if designed that way, which is what NS wants
  - o Complexity: This could be moderately complex. Data analysis needs to be disaggregated to compare different sub-groups, the CMO statements and analysis around those can be difficult. A strong understanding of hypotheses and program theory needed. Less complicated than large quantitative studies
  - o Cost: This is less costly than doing large quantitative surveys, and will require fewer volunteers
- Qualitative Comparison Analysis:
  - o Capacity: will need members of an M&E team or NS staff that have a very strong understanding of the program to determine outcomes and create factors. Quantitative data from previous research or monitoring can be used to find cases, so someone will be needed an understanding of basic statistics. Not a huge amount of capacity needed, but it is still complex.
  - o Preference: From interviews, people wanted qual and quant data collected and they wanted to look at beneficiaries and non-beneficiaries. This methodology could use non-beneficiaries in the comparison group, although that is not usual (why it didn't get a 5)
  - o Complexity: Finding the cases may be slightly complex, as it requires looking at a few factors, but the sample size is relatively small and requires fewer volunteers. The analysis will be predominately qualitative analysis, but there are scoring factors that can quantify the qualitative data, which is a bit more complex than regular qualitative data analysis.
  - o Cost: This should be fairly low cost. The sample size is much smaller, which means that fewer volunteers will be needed.
- Case Study:
  - o Capacity: intensive data gathering required, so capacity is required for that. Need to be able to collate all of the information gathered and then create questionnaires based off of what is left to be done, so need a strong M&E capacity to do that. Not as capacity required as some of the more involved studies. Basic data analysis skills needed for the gathered data.
  - o Preference: FbF team members wanted quantitative data gathered as well as qualitative, and liked the comparison between beneficiary and non-beneficiary, so doesn't fulfil all of the preference requirements.
  - o Complexity: This approach is moderately complex, as it requires the synthesis and triangulation of a lot of data and a detailed report, but not as complex as some other approaches
  - o Cost: Cost should be relatively low, because sample size is small and fewer volunteers are needed
- SCM:

- Capacity: There is moderate capacity needs for this methodology. Need basic quantitative and qualitative data analysis skills, and someone with M&E experience to be able to design questionnaires in a way that will identify successes
- Preference: This satisfies the desire of interviewees to have quantitative surveys and qualitative data to really identify what is working well. It does have a comparison, but not with non-beneficiaries, which is why it doesn't fit all preferences and get a 5. Could be modified...
- Complexity: This is a fairly straightforward technique. A success needs to be defined, then quantitative surveys are done to find successes and failures, then interviews to probe both. Analysis is also fairly straightforward. Research has stated that this is a quick way to gather good data
- Cost: Because a larger sample size is needed for surveys, there is a slightly higher cost involved than smaller sample size options, but most research has stated that it is generally cheap compared to other survey designs
- Collaborative Outcomes Reporting
  - Capacity: There are a few steps in this process that require capacity from the staff to be able to do well. These include the scoping inception and potentially developing workshops for the scoping part. The data trawl, where data needs to be found and analyze could require statistical capacity. Design of questionnaires and sample as well as data analysis will also require capacity. So will designing the final workshops and putting them on. So quite a bit of capacity is required.
  - Preference: This will provide qualitative and should provide quantitative data, both of which the programs want. It does not collect beneficiary and non-beneficiary info
  - Complexity: There are many steps involved in this process, none of which are straightforward. Therefore it is moderate-high in complexity.
  - Cost: It is hard to determine cost for this one, but it is assumed that data collection from volunteers, as well as putting on potentially two workshops (beginning and end) will be more expensive than some other methodologies
- Citizen Report Card:
  - Capacity: FGD needs to be run before the start, as well as sampling and survey development. Moderately complex
  - Preference: No qualitative data or comparison group, so not really what programs have said they wanted
  - Complexity: It is a fairly straightforward survey, the most complex part will be sampling and analysis. But with no comparison, sampling and analysis should be relatively simple
  - Cost: Large survey, which means the potential to use quite a few volunteers, which could be costly. There is also the suggestion to disseminate the information gained widely, which could potentially cost money
- Participatory Impact Assessment:
  - Capacity: Need a strong M&E team with deep knowledge into different interview and FGD techniques. The goal is to use FGDs and interviews to quantify qualitative data and do an analysis, so need both strong quant and qual analysis skills as well as very well-designed questionnaires and FGD

facilitation. While capacity for qual data needed and a bit of quant data, nothing prohibitive.

- Preference: Gives a comparison between beneficiaries and non-beneficiaries and qualitative data that interviewees have wanted. Does not give the overall quantitative data, so it is not everything that was wanted, but does attempt to quantify some of the qualitative data
- Complexity: Moderately complex with multiple interviews and FGDs. The FGDs analysis of the FGDs is also complex, as they attempt to create a baseline and repeated measures, then quantify that data
- Cost: moderate cost, because sampling % means that there could be large numbers of interviews required, meaning many volunteers.
- MAPP
  - Capacity: moderate capacity is needed for this methodology, as FGDs need to be facilitated to result in the specific info needed. Strong knowledge is needed on the methodology and FGD facilitation. Analysis is also not straightforward, will need good skills. Usually done by a consultant.
  - Preference: Is collects qualitative data and compares beneficiary to non-beneficiary, but does not collect much survey data.
  - Complexity: Not very complex, besides the need to set up the workshops and design the FGDs. While this is challenging, it is less complex than other studies.
  - Cost: This depends on the sample size and whether a consultant is hired. If a consultant is hired, then this could be fairly costly.
- Beneficiary Assessment:
  - Capacity: Need someone really well trained in participant observation, orientation is required, and a lot of research is required before starting data gathering. This requires moderate capacity. Also need someone to do qualitative analysis.
  - Preference: No quantitative data is gathered and there is no comparison between beneficiaries and non-beneficiaries
  - Complexity: Not very complex, besides the pre-data collection phase.
  - Cost: This should not be very costly, as not many people are required for data collection or analysis
- Most Significant Change:
  - Capacity: Not as much capacity is needed for this methodology. Sampling needs to be done, and there need to be volunteers or staff members that know how to conduct qualitative interviews.
  - Preference: Does not compare beneficiaries and non-beneficiaries, and does not collect quantitative data
  - Complexity: Is it not a complex process, the biggest complexity comes from gathering stakeholders at multiple levels to pick the most successful cases
  - Cost: It has been indicated that the first round of data collection could be from 100s of people. If that sample size is chosen, then it could be costly to have that many volunteers being used for longer interviews
- QUIP:
  - Capacity: the questionnaire development and sampling, as well as data collection, should be done by an outside evaluator. FbF team and M&E need little capacity for this.

- Preference: Does not collect quantitative data and does not compare beneficiaries and non-beneficiaries. Interviewees also mentioned wanting to be involved and build capacity, which this methodology does not really require
- Complexity: Fairly straightforward for FbF teams. They receive all the data and then just need to do qualitative analysis.
- Cost: Could be expensive to hire outside evaluators
- GEM:
  - Capacity: This requires some capacity upfront to go through all existing data and identify all possible explanations of what happened. Good M&E experience and data analysis skills needed. Questionnaires then need to be based off this info, so need to have strong questionnaire development skills.
  - Preference: Does not collect quantitative data and does not compare beneficiaries and non-beneficiaries
  - Complexity: Gathering all of the data needed to determine all possible explanations could be quite complex, as could determining who needs to be interviewed for what purposes
  - Cost: fewer interviews are needed, therefore fewer volunteers. This would not be very cost-intensive
- Outcome Harvesting:
  - Capacity: need good M&E and research capacity to comb through existing evidence and create outcome descriptions from that evidence. Also need to design learning questions and questionnaires as well as data analysis skills.
  - Preference: Does not collect quantitative data and does not compare beneficiaries and non-beneficiaries
  - Complexity: Not exceedingly complex. The hardest part is determining the outcome descriptions. It then just involves conducting interviews or FGDs and analyzing data in the context of the outcome descriptions
  - Cost: Not very costly, as fewer volunteers are needed for interviews and staff members can facilitate FGDs.

## **Appendix F: Success Case Method In-Depth Specifics for FbF Studies**

### **1. Component 1: Developing impact model and model of success**

There are two parts of the first component that must occur prior to data collection. The first is to develop an impact model. This will vary between each FbF program, because each program has been established using different hazards and early actions. Therefore, each program will have a slightly different theory of change to base the impact model on. This paper will use Bangladesh as an example.

#### ***Developing Impact Model***

Their theory of change is that by providing cash grants prior to flooding, then beneficiaries will be better able to evacuate, make fewer destitution sales, accrue fewer debts, consume more and better quality food, experience less psychological stress, suffer less disease, and resume productive activities than if they had not received the cash. Using this theory of change the impact model can be created. This can be done by the M&E team at different NS, with or without input from stakeholders. Even if stakeholders are not consulted, the draft impact model should be circulated amongst the assessment team staff and the FbF team to make sure everyone is in agreement.

#### ***- Example of Impact Model***

The most basic form of the impact statement is the following: By providing early actions before a forecasted hazard, beneficiaries should be able to prevent or reduce negative impacts such as loss of life and livelihoods. For the Bangladesh FbF assessment, the impact model would be more specific, and would constitute the following: By providing cash grants in advance of forecasted flooding, beneficiaries will be able to evacuate the affected areas, limit the number of destitution sales, consume good quality food, and generally prevent the negative impacts of flooding.

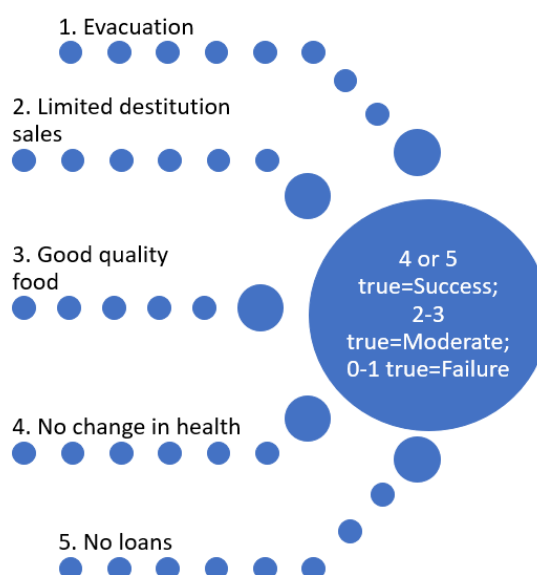
#### ***Creating Model of Success***

In most SCM cases, because it was initially designed as a way to determine whether interventions had an impact on organizational results such as ROI, the model of success is that participants that underwent training were able to apply these trainings successfully and produce better results. In a modified SCM study that assessed the impact of a nonprofit program, success cases have been participants that retained stable employment and housing. Other examples include having success be measured in high changes in psycho-social scores and having positive health and housing outcomes as judged by social care workers. This shows that the methodology can be adapted to different fields and models of success can vary greatly. All involve taking impact models and deciding which variables define a successful impact. For FbF programs the success model will vary by program,

because there are different combinations of hazards, early actions, and impacts. Coryn et al. (2009) created a success model using three different variables. If a participant had a positive outcome with all three variables, they were deemed a success. A similar technique will be used for FbF programs to define success.

#### - *Example of Model of Success*

To use Bangladesh as an example, the impact model involved several variables: evacuation, limiting destitution sales, consuming good quality food, experiencing no change in health, and not going into debt. Therefore a model of success will include these variables. A success for evacuation will mean that beneficiary households were able to evacuate prior to flooding with their families, livestock, and assets. A success in limiting destitution sales will mean that beneficiaries did not have to sell assets such as furniture, cookstoves, or livestock in exchange for money or food. Success in consuming good quality food will mean that the number, size, and variety of meals did not change from prior to the flooding. Success in terms of family health would mean no increase in health issues from prior to the flood. For the debt indicator, success would entail not needing to take out new loans after the floods. To be a success case, a participant will need to have had a success in at least four out of the five criteria. A participant will be classified a moderate case if they have had a success in two or three criteria. A participant will be classified a failure case if they have had a success in one or fewer of the criteria. The diagram below visualizes the breakdown



This can be adapted based on which indicators are being assessed and how many indicators are involved in the impact assessment.

## 2. Component 2: Quantitative Survey



The quantitative survey will gather basic information on the intervention and impact to find success and failure cases. Using the Bangladesh example, the survey will therefore need to ask questions about conditions prior to receiving the grant, what was done with the grant, and conditions after the flood in relation to health, livelihoods, and assets. This will help establish a baseline and then look at what changed. In the example of Bangladesh, questions need to provide enough information to assess whether a case was a success in any of the five categories. Previous uses of this methodology have used very simplistic surveys just to determine the success cases, which is one of the reasons that it has been described as a simplistic methodology. For FbF studies, the survey will be a bit more detailed, so that aside from determining success cases some quantitative data can be gathered from beneficiaries.

### *Sampling and Sample Size*

Many of the studies that have previously used this methodology have administered the survey to all participants, because they have had a small number of participants (Coryn et al., 2009; Clinton et al., 2007). The FbF programs have much larger number of participants, for example 1059 people were given the cash grants in Bangladesh (Gros et al., 2019). Therefore, a sample of participants needs to be taken. Sampling can be done in several ways, depending on the geographical area where the trigger occurs. If the area is small, one stage random sampling can occur. If the geographical area is larger, it may be better to do two-stage cluster sampling, where districts are first randomly sampled, and then individual beneficiaries are randomly selected from the sampled districts (USAID, 2021). In previous studies used to assess WFP interventions on nutrition, they have found that if there is limited data on the variables that are being assessed, a sample size of 150 to 250 beneficiaries should be sufficient. If the sampling design involves clustering, then 10 households per cluster should be sufficient, so between 15-25 clusters should be randomly selected (WFP, 2009). According to Israel (1992), with a population size of 1000 (which is similar to the Bangladesh intervention population) a sample size of 91 would result in an error of  $\pm 10\%$  and a sample size of 286 would result in an error of  $\pm 5\%$ . The error would be approximately  $\pm 5\%$  if the upper bounds of the WFP sample size is used. If the site has the capacity to conduct sample size calculations, that can be done in lieu of using the 150-250 sample.

### *Administration of Surveys*

Interviewees all mentioned that volunteers had a lot of experience in administering surveys, so they should be used. Administration of this survey will require the day long training, similar to previous surveys. In past IAs, there have been issues of not being able to find beneficiaries, so backups will need to be identified in advance.

### *Analysis*

After the surveys are administered, the data will need to be analyzed to find the success and failure cases. This will require someone with basic knowledge of

statistics to look at the data. In prior studies, there have been specific cut offs for numbers that have been considered a success and failure, and all other cases are classified as moderate (Coryn et al., 2009). For the example of Bangladesh, analysis will need to be done on whether a beneficiary was a success in the five variables that were established in the model of success. This survey data should also be used to provide quantitative data on the intervention, as well as demographic characteristics. This survey can provide information on what beneficiaries experienced before, during, and after the flooding as well as how they used the grants.

### **3. Component 3: Qualitative Survey**

The next step of this methodology is to sample a few of the successes and failures to interview to determine what factors led to the success and what led to the failures. Different studies have done this differently, with some choosing to only look at successes, but for FbF it is important to have a comparison with failures and to see why things didn't go well (Brinkerhoff, 2005).

#### *Sampling and Sample Size*

The number of successes and failures has varied, but there are usually between 2 and 6 people from both success and failure groups interviewed (Coryn et al., 2009; Clinton et al., 2007). As FbF studies have larger beneficiary numbers than previous studies done, the sample size should be a bit larger to help understand different reasons for success and failure, between 6 and 10 in each group. Literature on qualitative sample sizes has suggested that a sample of 10 participants is sufficient if participants hold a large amount of information relevant to the study. All participants will be able to provide large amounts of information, therefore the sample size of 10 seems sufficient (Multerud et al., 2015). Other research has suggested that 15-30 is the optimal sample size for single case interview studies (Marshall et al., 2013). The upper range of the sample size proposed here fits these requirements.

#### *Questionnaire Development and Administration*

Qualitative questionnaires will be developed differently for successes and failures. The goal of these questions will be to allow an open-ended conversation about why successes were successful and why failures failed. In the example of Bangladesh, this would include questions on how the grant money was spent, what challenges occurred when preparing for floods, and what occurred to the beneficiary after the flooding (Gros et al., 2019). These questions should be developed to probe for specific information on successes and failures so that FbF staff can get a strong understanding of which actions worked, why they worked, and if there are specific characteristics or actions that were taken that have led to successes. For failures, information needs to be collected to determine if some actions didn't work and why they didn't work. If actions didn't work, the interviewee should be probed to determine why they didn't work. If some actions that were taken with the cash were not successful, this information needs to be determined in the interviews so that

recommendations for actions can be given prior to cash transfers. Questionnaires will be administered as semi-structured interviews. Staff members or volunteers that are trained in qualitative interviews can conduct the interviews.

#### **4. Component 4: Analysis and Reporting**

Qualitative analysis will focus on the reasons that success cases became successes and why failures failed. This will help determine if there is a pattern to success and failure, and what some of those conditions are (Piggot-Irvine et al., 2009). All of the results are then written up into a report. Most of these reports are typically written as “success stories,” but in the case of FbF, the descriptive statistics and the failure stories should also be included to provide context and detailed analysis of failures (Brinkerhoff, 2005). Looking at failures will hopefully provide evidence of changes that can be made to improve the interventions. The final report should include quantitative data on the intervention and participant characteristics as well as in-depth success and failure stories.

#### **5. Modifications**

Depending on the capacity and funding available at different National Societies, there are several modifications that can be made.

##### ***Adding non-beneficiary groups***

First, many interviewees mentioned the desire to look at non-beneficiaries in addition to beneficiaries. If the capacity and funding is available, non-beneficiaries could be sampled to take the survey and several success and failure cases should be interviewed. This will give a brief picture of the characteristics of the non-beneficiaries and some of their success and failure factors. There are several different sampling techniques and sample sizes that could be used depending on the capacity of the NS and the funding that is available.

##### ***1. Include non-beneficiary success and failure interviews***

This modification would include purposive sampling of a small sample of non-beneficiary failure and success cases. This will involve using the same variables to determine the beneficiary success model, and sampling people that have had a successful preparation for the flood based on those variables and those that have not been successful. Information for sampling can come from either stakeholder knowledge or post-disaster surveys done by NS. The sample size should be similar to the beneficiary interviews, between 6-8 participants in each group. Previous research on sample sizes for qualitative research has been inconclusive, but a meta-analysis suggested that a sample size between 15 and 30 should be sufficient, so combined with beneficiaries this sample size is sufficient (Marshall et al., 2013). They will be asked similar questions to the beneficiaries. The data collected from the interviews will be able to establish how people that did not receive funds responded to the flood, and different reasons that they had successes or failures in response. While it does not provide a representative sample of the non-beneficiary population,

if the NS does not have greater capacity it will at least provide non-beneficiary information to draw some conclusions about the program impact. The figure below provides an overview of how this modification would work.

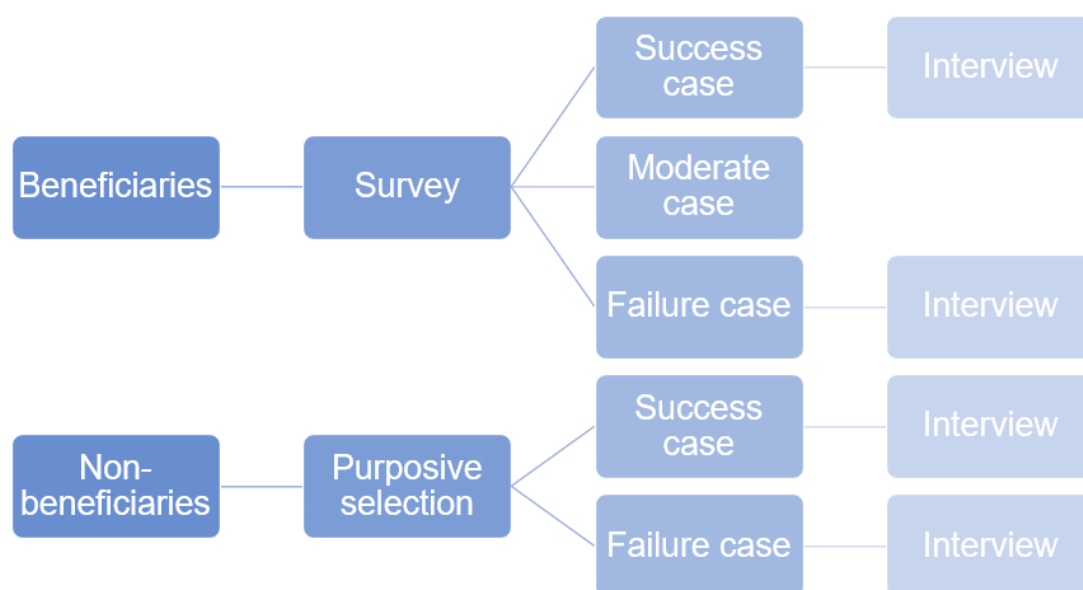
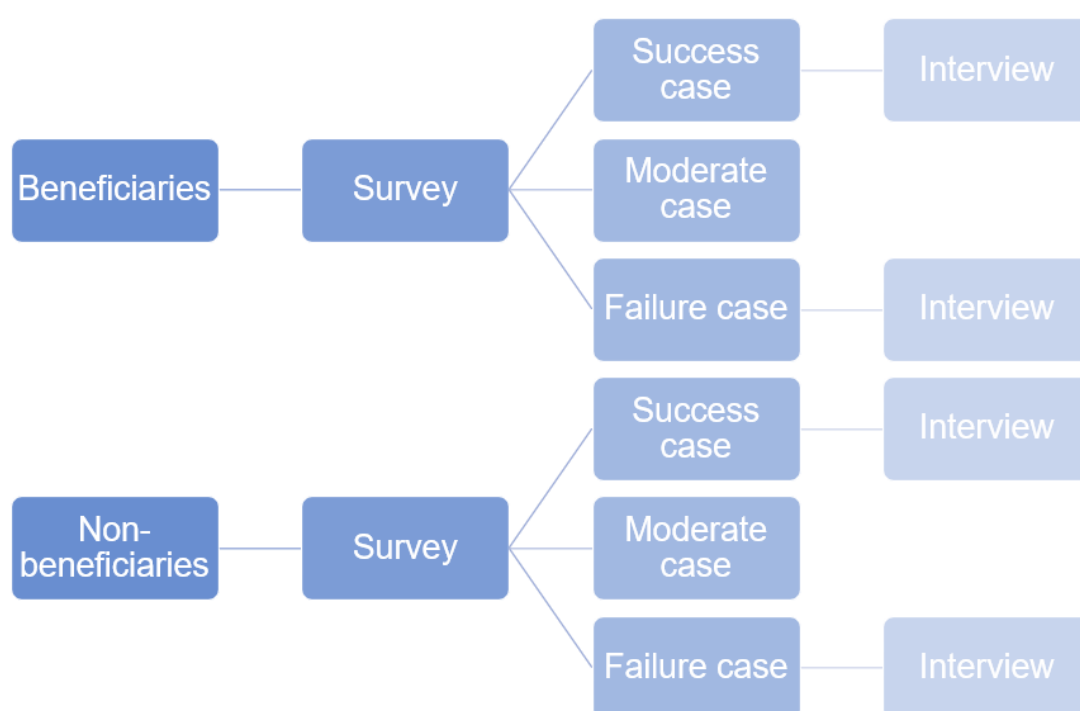


Figure: Modification with non-beneficiary interviews

## 2. Include non-beneficiary survey and interviews

This modification is most appropriate for NS that have a strong statistical and M&E capacity. It involves surveying both beneficiaries and non-beneficiaries, and then finding successes and failures from both groups and interviewing them to get in-depth information on both groups. There are different sampling techniques that can be used to determine sample size for this modification. If the data is available, and the NS has the capacity to do so, the traditional sample size calculation to compare groups should be used. If those sampling techniques are too difficult for NS to do, or the NS is not able to conduct a survey with the large sample sizes that will result from sample size calculations, then the WFP sampling technique can be used. To determine the impact of nutrition programs, they use the sample size of 150 to 250 per group as a rule of thumb if not enough data is available to calculate sample sizes. Using that rule of thumb, 150 to 250 beneficiaries and non-beneficiaries should be surveyed (WFP, 2009). Fewer participants could be surveyed, but that would limit the representativeness and increase sampling error. It has been suggested a minimum of 100 participants per group should be sampled (Singh & Masuku, 2014). The downside to this sample size is that it does not have the same rigor as determining the sample size statistically. If the capacity and resources are available, sample size calculations could be done. Sampling technique will be the same as the SCM detailed above, with either simple random sampling or cluster random sampling used.

Once the beneficiaries and non-beneficiaries are surveyed, that information will be used to determine who in both groups is a success and failure. A sample of the successes and failures will be interviewed to determine what factors led to these successes and failures. This modification allows for the comparison between both the beneficiary and non-beneficiary survey and the success and failure interviews. This should give the best overview of what happened during the intervention and what the impact was. It should also provide evidence on why and how people successfully used the intervention, along with factors that can lead to success even without the intervention. Below is a diagram that explains the events in the assessment.



*Figure: Modification with non-beneficiary survey and interviews*

### **Adding longitudinal data**

One study that used a modified version of SCM decided to include a time series element to the design, re-surveying participants to see if participants moved categories at three different time points (Coryn et al., 2009). This could be a modification to FbF impact assessments if a program had the capacity and wanted more long-term data on impacts.

At each time point, the survey and interviews will need to be conducted, greatly increasing cost and capacity required. Despite this, this modification provides information on long term impacts, as well as robust data on participants that change groups (success to moderate/failure or failure to moderate/success).