

# MEASURING RESILIENCE

## Lessons learned from measuring resilience in Oxfam's large-N Effectiveness Reviews



Photo credit: Alejandro Chaskielberg/Oxfam

**Resilience has rapidly become one of the most prominent objectives for the development sector, so ascertaining how best to measure it is an essential task for practitioners working in monitoring and evaluation. In this discussion paper, we outline the main insights emerging from the series of large-N Effectiveness Reviews, a set of quantitative studies that aim to evaluate impact and generate learning from a random sample of Oxfam's projects. We also consider how this measurement approach may adapt as ideas about resilience change both within Oxfam and in the development sector at large.**

### OXFAM GB

#### **Oxfam Discussion Papers**

Oxfam Discussion Papers are written to contribute to public debate and to invite feedback on development and humanitarian policy issues. They are 'work in progress' documents, and do not necessarily constitute final publications or reflect Oxfam policy positions. The views and recommendations expressed are those of the author and not necessarily those of Oxfam.

For more information, or to comment on this paper, email [ppat@oxfam.org.uk](mailto:ppat@oxfam.org.uk)



# CONTENTS

<b>Contents</b> .....	<b>2</b>
<b>1 Introduction</b> .....	<b>3</b>
<b>2 Measuring resilience in Effectiveness Reviews</b> .....	<b>4</b>
<b>3 Disentangling the causal chain</b> .....	<b>5</b>
<b>3.1 Measuring resilience in cases where shocks and stresses have already occurred</b> .....	<b>5</b>
<b>3.2 Separating elements along the causal chain</b> .....	<b>6</b>
<b>3.3 Risk of a circular definition</b> .....	<b>7</b>
<b>4 A multidimensional approach</b> .....	<b>8</b>
<b>4.1 Dimensions of resilience</b> .....	<b>8</b>
<b>4.2 Aggregation and defining thresholds</b> .....	<b>9</b>
<b>5 Generating indicators and practicalities</b> .....	<b>10</b>
<b>5.1 Resilience to what type of shock, stress or uncertainty?</b> .....	<b>10</b>
<b>5.2 Tools for guiding discussions about indicators of resilience</b> .....	<b>11</b>
<b>5.3 Measuring resilience characteristics without a comparison group</b> .....	<b>12</b>
<b>5.4 Should diversification be an indicator of resilience?</b> .....	<b>12</b>
<b>5.5 Indicators of innovation potential</b> .....	<b>13</b>
<b>6 Perception-based measures</b> .....	<b>14</b>
<b>6.1 Perception-based measures for specific characteristics of resilience</b> .....	<b>14</b>
<b>6.2 Perception-based measures of overall resilience</b> .....	<b>15</b>
<b>7 Conclusions and next steps</b> .....	<b>16</b>
<b>Notes</b> .....	<b>18</b>
<b>Bibliography</b> .....	<b>19</b>

# 1 INTRODUCTION

Resilience has rapidly become one of the most prominent objectives for the development sector. In part, this has been driven by the recognition that humanitarian and longer-term development work could and should be better coordinated, and that resilience provides an overarching framework to help integrate these two types of activity (Frankenberger, et al., 2014). This point is sharpened because uncertainty, both in terms of economic volatility and the frequency and intensity of weather and climate-change related shocks, is rising (Schipper & Langston, 2015). It is no surprise, therefore, that support for resilience-building initiatives has attracted substantial funding and resources. For example, Béné et al. (2015) estimate that the total global spend on resilience programming exceeds \$5bn.

Although there is still some variation among development practitioners about how to define resilience, the concept is gradually becoming more consistent throughout the sector (Béné, et al., 2015). However, in spite of these efforts to clarify the definition of resilience, there is still much work to be done to develop a coherent framework for how it should be measured. Having identified resilience as a key outcome area for the organisation, Oxfam GB has been grappling with this measurement challenge for several years now, and we are keen to share lessons learned from our experience.

Oxfam Great Britain (GB) operates in more than 50 countries, across more than 300 different programmes, within which sit over 1,000 projects that are designed to help end poverty. In order to understand the impact of this work, mature and closing projects are randomly selected under 6 key outcome areas each year and rigorously evaluated as part of Oxfam GB's Global Performance Framework. These intensive evaluation processes are known as *Effectiveness Reviews*, and they consider whether there is a cause–effect relationship between the intervention and observed outcomes/impact, even where the intervention may be only one contribution to that change. It is critical that the changes we are seeking to achieve are properly measured, and Oxfam has invested in developing appropriate evaluation designs and measurement approaches for the outcome areas, many of which are considered hard-to-measure benefits (HTMB) (Hutchings, 2014).

Effectiveness Reviews of community-level development programmes working to build resilience employ a quasi-experimental design to consider the counterfactual, collecting information through household surveys and other means from both intervention and comparison populations.<sup>1</sup> Advanced statistical methods, such as propensity score matching and multivariate regression are used to reduce bias in this comparison by controlling for any measured differences existing between the two groups. In order to define and measure resilience, Oxfam has developed a multidimensional framework to help conceptualise and structure the drivers and characteristics of resilience. Further details on the Effectiveness Reviews can be found on Oxfam's Policy and Practice website (<http://policy-practice.oxfam.org.uk/>).

The purpose of this discussion paper is to outline the main lessons learned about measuring resilience from the set of large-N Effectiveness Reviews. We are seeking to make a contribution to the sector by sharing Oxfam's ongoing learning and experience to date, but also hope to stimulate discussion, wider reflection and feedback by considering those areas we are seeking to strengthen and develop.

We start in Section 2 by describing the general method Oxfam has used for conducting Effectiveness Reviews on projects aimed at building resilience, focusing in particular on the causal chain that leads from Oxfam's projects to development outcomes. In Section 3, we consider the difficulties that arise in trying to disentangle this causal chain. In Section 4 we examine how the multidimensionality of Oxfam's approach to measuring resilience has been borne out in the Effectiveness Reviews. In Section 5, we outline some of the practical issues we

have experienced in actually generating indicators from the multidimensional conceptual framework, and in Section 6, we focus in particular on the validity of self-perception-based measures. Finally, in Section 7, we conclude and consider how Oxfam’s framework for measuring resilience may adapt in the future to reflect changes in the way resilience is understood both within in Oxfam and in the development sector at large.

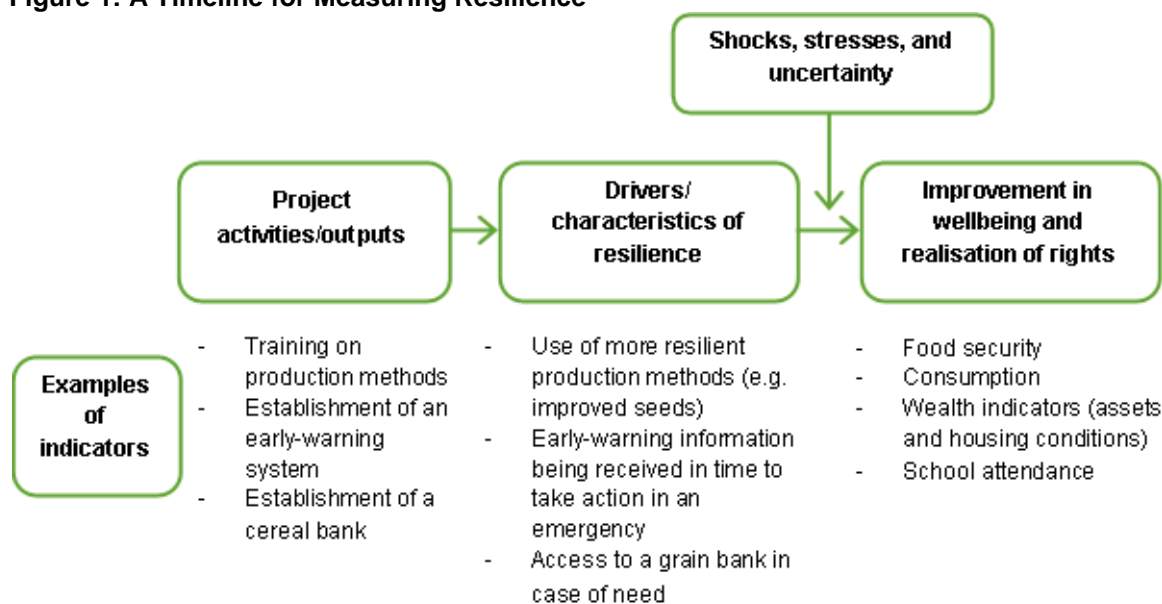
## 2 MEASURING RESILIENCE IN EFFECTIVENESS REVIEWS

Oxfam’s definition of resilience is ‘the ability of women and men to realise their rights and improve their wellbeing despite shocks, stresses and uncertainty’. As described in the Oxfam Framework and Guidelines for Resilience Programming, ‘Our focus is on people’s agency. We strengthen capacities that enable people to positively adjust to shocks, changes and uncertainty and to make choices that shape more resilient futures.’ (Jeans, et al., 2015). Oxfam’s report, *No Accident: Resilience and the Inequality of Risk* suggests that resilience means people have ‘hope for the future and real choices about how to live their lives and adapt to change. Our ambition goes beyond helping people survive one shock after another; we want people to thrive despite shocks, stresses and uncertainty.’ (Hiller & Castillo, 2015).

In line with this definition and discussion, the final ‘impact’ we hope to see from resilience-building programmes is an improvement in the realisation of rights and in wellbeing. If a person, household or community has been able to realise their rights and improve their wellbeing over time, even while suffering shocks, stresses and uncertainty, then we can describe that person, household, or community as being resilient. In that case, the issue of measuring resilience itself is straightforward: we can simply observe the improvement in wellbeing and/or the realisation of rights. However, in our Effectiveness Reviews – and in other evaluations – we are usually seeking to evaluate the impact of resilience-building work only shortly after that work was carried out (at most 12 to 18 months after the end of a project). In most cases we would not expect to see increased resilience leading to improvements in wellbeing over such a short time. Instead, we try to identify drivers of resilience: characteristics that we believe are important for people’s ability to thrive despite shocks, stresses and uncertainty. We then create indicators and survey questions related to each of these characteristics.

The diagram shown in Figure 1 helps to visualise this. We carry out project activities that are intended to have a positive effect on the drivers or characteristics of resilience. Over time, people will experience shocks, stresses and uncertainty, which will affect their wellbeing outcomes and their realisation of rights. In principle, we could wait and look only at the wellbeing outcomes in order to assess the impact of the project activities. However, within the timescale of our evaluations, we normally need to look at intermediate steps along this causal chain, which means looking at drivers or characteristics of resilience.<sup>2</sup>

Figure 1: A Timeline for Measuring Resilience



## 3 DISENTANGLING THE CAUSAL CHAIN

Although the diagram shown in Figure 1 provides a useful starting point, there are at least three reasons why disentangling the elements along this causal chain may be more challenging in practice. Firstly, there may, in fact, be cases where resilience must be measured *after* shocks, stresses, or uncertainty have already occurred. Secondly, it may be tricky to separate drivers or characteristics of resilience from final outcomes, that is, those outcomes reflecting realisation of rights and improvements in wellbeing. Thirdly, it is often unclear how to distinguish the direct outputs of a project from the drivers or characteristics of resilience. As we explain below, there is a risk of building into our measurement framework an assumption that resilience has increased simply by virtue of a project having delivered on its planned outputs. This can lead to resilience being given a circular definition.

### 3.1 MEASURING RESILIENCE IN CASES WHERE SHOCKS AND STRESSES HAVE ALREADY OCCURRED

Some of the contexts in which we are trying to evaluate projects' impact on resilience are subject to continuing climate-related stress, particularly in the Sahel and other arid regions.<sup>3</sup> In these regions people are subject to weather shocks and stresses almost every year. This means that we can observe the final outcomes (the realisation of rights or an improvement in wellbeing) directly, by looking at indicators of food security, nutrition, asset ownership, school attendance and so on. In these scenarios, it has been suggested that we do not need to measure the intermediate outcomes (the drivers/characteristics of resilience) at all.

There are, however, a number of reasons to include drivers or characteristics of resilience alongside final outcomes, even for the evaluations of projects in areas that are subject to continuing climate-related stress. In particular, if we are being asked to evaluate projects' effects on resilience *now*, it is not sufficient simply to look at final outcome measures as an indicator of how people (or households or communities) have dealt with shocks, stresses and uncertainty in the *past*. It may be that people have been able to cope with past shocks through selling their assets (e.g. pastoralists selling off their livestock), but that this means they are now more vulnerable to future shocks. The concept of resilience implies something about coping with shocks, stresses, and uncertainty in the *future*, so using indicators that capture the drivers or characteristics of resilience is necessary, even in cases where shocks are frequent.

Incorporating indicators of drivers or characteristics is even more important in places where resilience-building means enabling people to deal with the effects of some *infrequent* shock that may not have occurred by the time the evaluation is carried out.<sup>4</sup> In these contexts, final outcomes alone will provide a very noisy signal of a project's resilience.

In practice, in recent Effectiveness Reviews we have drawn a pragmatic balance between collecting indicators of final outcomes (mainly food consumption, food security, and wealth indicators) and indicators of the drivers or characteristics of future resilience.<sup>5</sup> In these cases, the set of indicators of the drivers or characteristics of resilience has been reduced to allow us to collect more detailed data on final outcomes. The main limiting factor has been the number of questions we can ask in a one-hour survey.

## 3.2 SEPARATING ELEMENTS ALONG THE CAUSAL CHAIN

With the causal chain in Figure 1 in mind, the question remains over how to use the menu of indicators of direct project outputs, drivers or characteristics of resilience, and improvements in wellbeing or realisation of rights, to create an aggregate measure of resilience. One approach may be to construct a composite index from these indicators. For example, our aggregate resilience measure could include characteristics such as whether the community is covered by an early-warning system (which may be a direct output of the project), whether the household has a diversified livelihood, savings, and social support (which may be considered as drivers or characteristics of resilience), and also the household's actual experience of food security in the period prior to the survey (that is, a final wellbeing outcome). However, given the causal links between each of these components, this approach would make interpreting the overall index of resilience somewhat confusing.

We prefer to restrict the indicators of resilience to those that are at an intermediate stage between direct project outputs and final outcomes. For example, in the case of the early-warning system, indicators of resilience would not simply focus on whether the early-warning system was actually functioning, but rather would use higher-level outcomes, such as respondents' confidence in receiving the warning before a disaster, and whether or not they know what to do if they receive a warning. This type of reasoning enables us to distinguish the direct project outputs from the drivers or characteristics of resilience.

An additional complication is that some of our final outcome measures – such as food security and asset wealth – may also be understood as drivers of resilience. A person or a household that already has an adequate diet is more likely to be healthy, and is therefore probably better able to thrive in the face of future shocks. Similarly, owning assets is an important factor in being able to cope with future difficulties, either because assets can be used for production, or because they can be sold to generate income. It is therefore ambiguous whether factors such

as food security and asset wealth should be included as indicators of resilience as well as final outcome indicators in their own right.

In recent Effectiveness Reviews, we have not included households' food security situation in the resilience index, in part because finding robust indicators of food security for the evaluations is difficult. We have, however, usually included 'ownership of productive assets' as an indicator of resilience, typically measured by assessing households' ownership of agricultural equipment, vehicles, mobile phones, and various other context-relevant items. Admittedly, these decisions are to some extent arbitrary, and more work is required to hone these types of choices.

### 3.3 RISK OF A CIRCULAR DEFINITION

One important problem in measuring resilience arises when our measurement framework is built on the same assumptions as those used to design the project or programme being evaluated. This risks creating a self-referential definition of resilience. For example, we may believe that increasing households' crop diversity is important for building resilience, and so we implement a project aimed at increasing diversity of crops being grown. When we come to draw up a framework for measuring resilience in order to evaluate this project, we make the same assumption that crop diversity is important to resilience. Thus, if we find that the project has increased crop diversification, we will report that the project has successfully increased resilience, even though the link between crop diversification and resilience has not been tested.

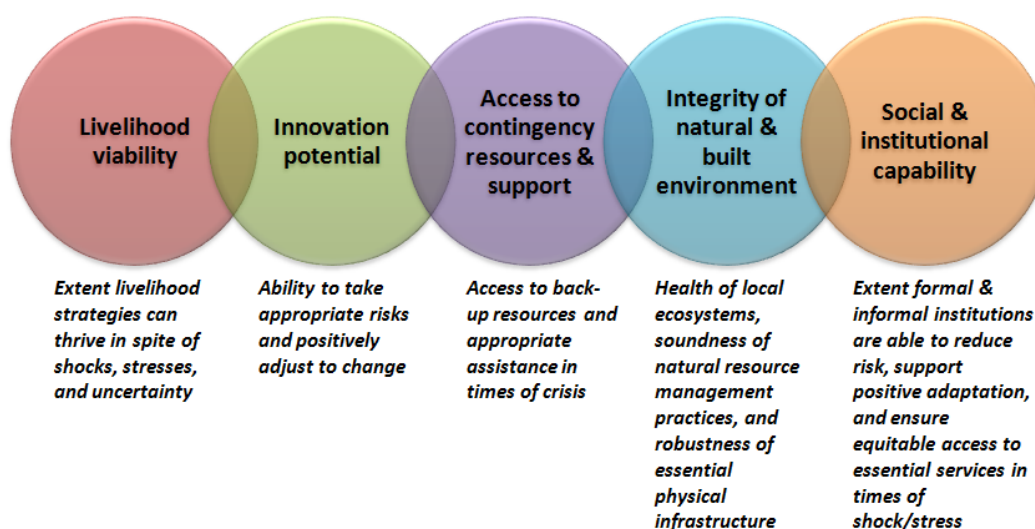
This should be less of a problem for *ex post* impact evaluations, like the large-N Effectiveness Reviews, than for *ex ante* evaluations, since we are defining characteristics of resilience some years after the project or programme was designed. This allows the ideas and assumptions about drivers of resilience to change between the design of the project and its evaluation. However, insofar as the assumptions about how direct project outputs can build resilience remain the same between the start and the end of the project, even *ex post* evaluations may risk defining resilience in a circular way. As such, the best response we have is to seek opportunities to test the assumptions about characteristics of resilience wherever possible. One way to do so would be to use follow-up surveys some time after carrying out *ex post* evaluations like the Effectiveness Reviews, to see whether the indicators of resilience that were identified as characteristics were associated over time with more positive outcomes.

# 4 A MULTIDIMENSIONAL APPROACH

## 4.1 DIMENSIONS OF RESILIENCE

To facilitate the timeline for measurement shown in Figure 1, we require a way of conceptualising and structuring all the potential indicators for drivers and characteristics of resilience. Oxfam has developed a 'multidimensional approach', which provides a framework within which to develop these indicators. The details of this approach are provided by Bushell and Hughes (2013), so we do not repeat the explanation here. Nonetheless, the dimensions are shown in Figure 2.

**Figure 2: Oxfam's Multidimensional Approach to Measuring Resilience**



Clearly these dimensions all overlap to some extent, such that some indicators could be categorised under two or more dimensions. For example, ownership of livestock may be understood as both a productive asset and therefore a component of 'Livelihood Viability', or as a store of wealth, fitting under 'Contingency Resources and Support Access'. However, the overlaps do not necessarily correspond to the position of the 'bubbles' in Figure 2. Indeed, alternative versions of this framework have placed 'Social & Institutional Capability' at the centre, given that this is fundamental to building resilience in any of the other dimensions. Understanding these areas of overlap is particularly important for the aggregation of indicators into an index of resilience, as we discuss in Section 4.2.

In the Effectiveness Reviews, the model of the five dimensions has mainly been useful as a prompt to ensure that we consider a broad range of factors when searching for indicators of resilience. For example, some discussions in communities about indicators of resilience have tended to focus on the resilience of household-specific livelihood activities, with characteristics of 'Innovation Potential' or 'Social & Institutional Capability' only being mentioned after further prompting. However, it has proved beneficial to present this framework towards the end of the discussion about characteristics of resilience, so that suggestions for drivers or characteristics of resilience are not too heavily influenced by the examples of characteristics that have been used in previous Effectiveness Reviews.



As we discuss in Section 7, the use of the five dimensions is currently under review. We are developing a measurement strategy that better reflects the three capacities – absorptive, adaptive and transformative – that are increasingly being used in Oxfam’s programme design and across the rest of the development sector.

## 4.2 AGGREGATION AND DEFINING THRESHOLDS

In the Effectiveness Reviews, we convert each indicator of resilience into a binary measure, and then aggregate these to produce an index of resilience. Vitally, this allows us to present an overall picture of each project’s effect on resilience, while also capturing the important interactions between specific drivers and characteristics of resilience. However, as we explain below, there are at least three important considerations with aggregating indicators: (1) there are a number of possible aggregation methods; (2) there is little objective basis for assigning weights to dimensions and indicators; (3) converting indicators into binary variables requires applying some cut-offs or thresholds, which are difficult to establish and may cause us to throw away information.<sup>6</sup> Since these issues are partially unresolved, it is not clear that our aggregate measures are more revealing or informative than the individual characteristics. Thus, we have preferred to report the overall resilience index alongside its constituent indicators.

Firstly, for the latest wave of Effectiveness Reviews, we have preferred to keep the aggregation method as simple as possible, by focusing attention on what we call the ‘base resilience index’. This means the percentage of characteristics in which each household scores positively, against the binary thresholds we have defined. One alternative, which we also include in the reports, is to use the ‘Alkire-Foster resilience index’. Again, this is explained fully in Bushell and Hughes (2013).<sup>7</sup> The appeal of the Alkire-Foster resilience index is that it gives less weight to those households that score positively in many (normally two-thirds) of the characteristics of resilience, and so focuses the analysis on differences between those households that are less resilient. There is, however, currently little basis for us to decide on what the overall threshold used in the calculation of the Alkire-Foster index should be, and it can be complex to interpret. As such it has proved useful to report the base resilience index and the Alkire-Foster index together.

Secondly, aggregation requires some method for weighting each of the five dimensions and then also the indicators within each dimension. In some of the early Effectiveness Reviews, different weights were given to each of the five dimensions, and then indicators were weighted equally within each dimension.<sup>8</sup> More recently, however, we have assigned equal weight to each indicator.<sup>9</sup> In practice, this means that dimensions with more indicators are given extra weight compared to those with fewer indicators. This agnostic approach is justified by the fact that there is no obvious guidance that we can use to determine whether or not one indicator is more important than another. That said, there is some de facto weighting involved simply in deciding which indicators to include. If we weight the dimensions first (as in some of the earlier Effectiveness Reviews), adding characteristics under a particular dimension implicitly down-weights the influence of each indicator on that dimension and hence the overall resilience index. In addition, we may weight certain *households* more than others, if we add indicators that are only relevant to a subset of the population. For example, if we include access to veterinary care and vaccination of livestock as indicators of resilience as well as simply the ownership of livestock, then we are effectively giving greater weight to those households that own livestock at all.

Finally, aside from the question of whether we want to aggregate our resilience indicators, we have normally presented the results for each indicator in binary terms; that is, the proportion of households above or below the threshold for the relevant indicator. This makes the presentation and explanation of the results clearer. However, this presents a difficult question over the extent

to which the thresholds for each characteristic of resilience should be determined in advance by discussion with local staff and partners, and the extent to which the thresholds should be adapted once the survey data is available. In practice, it has been helpful to wait until seeing the survey data before setting the thresholds, to ensure that the resulting binary variables do not mask important differences between the intervention and comparison groups. Nonetheless, it is important to check the sensitivity of our results to alternative thresholds, and to examine alternative forms of each indicator – such as continuous variables for number of livestock, or categorical variables to capture household wealth.

## 5 GENERATING INDICATORS AND PRACTICALITIES

In this section, we illustrate some of the practical considerations involved in using the multidimensional approach to actually develop indicators for the drivers or characteristics of resilience. First, we describe the tension that arises when trying to evaluate projects that are often targeted at one type of shock, stress, or uncertainty with a broad notion of resilience in mind. Secondly, we consider some of the tools and techniques that have been used during focus groups and discussions with project staff during the fieldwork for Effectiveness Reviews, which help to generate indicators in a participatory way. Finally we outline two specific difficulties arising from the multidimensional approach: (1) How to interpret diversification; (2) How to identify indicators of Innovation Potential.

### 5.1 RESILIENCE TO WHAT TYPE OF SHOCK, STRESS OR UNCERTAINTY?

In principle, the multidimensional approach requires us to evaluate households' ability to deal with shocks and stresses and adapt to emerging trends and uncertainty in a general sense, without focusing on one particular type of potential shock. In practice the majority of the projects that are evaluated for their impact on resilience have been consciously focusing on one main type of shock – typically drought, flood or some other natural disaster.

Our process of identifying characteristics of resilience in each particular programme context has relied on discussions with local programme staff and partner organisations, and normally with people in local communities. During these discussions, we have tried to keep the thinking about shocks, stresses and uncertainty as broad as possible. It is not always easy to ask about characteristics that enable people to cope with shocks without talking about specific examples, such as droughts, floods, food price rises or medical crises in the family. However, the number of examples that can be discussed is clearly limited, and less thought and attention tend to be given to thinking of characteristics that relate to the later examples. Also, programme staff may be uncomfortable with evaluating resilience to types of shocks, stresses or uncertainty that were never intended to be addressed by the project. Again in cases like this, we have tried to strike a pragmatic balance between a broad understanding of resilience and addressing the specific objectives of the project.

## 5.2 TOOLS FOR GUIDING DISCUSSIONS ABOUT INDICATORS OF RESILIENCE

During conversations with project and partner staff, as well as in focus groups with local communities, it is useful to have tools and techniques to help guide the discussion. The best methodology that has emerged for doing this in the Effectiveness Reviews is adapted from the tool developed for focus groups in Niger by Venton (2013).<sup>10</sup> In particular, we have mainly used the following two questions, taken directly from Venton's tool:

1. What is the minimum level a household must be able to sustain at any time, to provide for its members? – i.e. what is the minimum level at which we would consider a household to be 'resilient'?
2. What are the factors that help to maintain this state/minimum level? What would give you confidence that you would be able to maintain this minimum level? What are the characteristics of households that are able to maintain this level?

The aim of the first question is to draw a threshold for what a 'resilient household' should look like, in terms of dietary diversity, number of meals per day, livestock holdings, ability to send children to school, or some other outcome. In general, we have tried to keep that as quite a short discussion. The second question then seeks to identify what the drivers or characteristics of resilience should be. This is normally quite an in-depth discussion, and the output should be a detailed list of drivers or characteristics of resilience.

The questions are focused at the household level since in the Effectiveness Reviews we make comparisons in outcomes between households. However, many of the characteristics that affect household-level resilience may operate at different scales. Firstly, intra-household factors may affect the overall resilience of the household. For example, we often observe that it is important that both men and women in the household have some diversification in their livelihood activities. Secondly, households are likely to be dependent on community- or institutional-level characteristics of resilience, especially if they utilise community support structures or social protection mechanisms.

One issue we have found with using the two questions outlined above is that there is not always a clear distinction between responses to the first and second questions, so that some elements that we would regard as drivers or characteristics of resilience are raised in the first question.<sup>11</sup> Perhaps, therefore, it would be beneficial to reformulate these questions to make them clearer for future Effectiveness Reviews. However, provided that the discussions generate functional new indicators, the blurring of the two questions may not be too problematic. We have also used a number of follow-up questions to help explore and clarify some of the points raised during these discussions.<sup>12</sup>

Some have argued that it may be possible to go further than simply eliciting a list of indicators from these discussions, in particular by asking project and partner staff as well as local communities to inform the weightings used for each dimension and each indicator (Holland, 2013; McGregor, et al., 2015). As we describe in Section 7, we are now experimenting with these types of methodologies for the current round of Effectiveness Reviews.

## 5.3 MEASURING RESILIENCE CHARACTERISTICS WITHOUT A COMPARISON GROUP

The large-N Effectiveness Reviews use a comparison group to ascertain what *would have* happened to project participants, had they not, in fact, participated in the project. Without a comparison group, it is very difficult to generate causal estimates for the effects of a project. We are, therefore, somewhat sceptical about the utility of collecting data on outcome indicators only for project participants and looking solely for changes over time without the use of a comparison group.

However, we should be especially careful when collecting data in this way on outcome indicators of resilience. Some of the indicators that we often use – such as diversification in livelihoods activities, the range of crops grown, or the types of livestock owned – are likely to be affected by year-to-year changes in weather conditions, especially in arid regions where monitoring or measuring resilience is particularly important. Given this annual variation, a comparison of resilience indicators across time (e.g. comparing baseline and endline data) may reflect people's responses to particular seasonal conditions more than reflecting changes in households' resilience.

At this stage we do not have a better suggestion for tracking indicators over time. Rather, we prefer not to invest in quantitative data collection at the household level without the inclusion of a comparison group. However, we recognise that this is often the minimum required for accountability purposes, even if it is not ideal for learning. In these circumstances, careful consideration should be given to identifying indicators that are not likely to reflect temporary year-to-year changes to the weather, asset prices, and other short-run variables.

## 5.4 SHOULD DIVERSIFICATION BE AN INDICATOR OF RESILIENCE?

There is often a tendency in programme work, and thus the discussions about indicators, to assume that more diversification is always positive for risk management. We have followed this in our Effectiveness Reviews by generally assuming that diversification of income sources constitutes an indicator of resilience. However, this assumption is open to question. There is a large literature outlining the factors that determine whether diversification will successfully build resilience, and here we outline the most pertinent issues from the perspective of Oxfam's Effectiveness Reviews.

Firstly, if households diversify their income sources, this may result in livelihood strategies with lower *variance* (less fluctuation over time), but also lower *expected returns* (lower returns on average) (Dercon, 2008). For example, if producers are encouraged to diversify the range of crops they grow, they may lose out on economies of scale that would otherwise be available in harvesting or marketing. Moreover, in Oxfam's Effectiveness Reviews, we have often asked households how many income earners they have. However, we should be cautious about assuming that households with more income earners are more resilient, particular if the extra earning power relies on children or other household members being taken out of education (Beegle, et al., 2006). Distinguishing positive adaptation from these types of 'mal-adaptation' is important for ensuring projects build both the short- and the long-run aspects of resilience (Magnan, 2013).

Secondly, there may be important limits to how much households are actually able to diversify. This results in part because much of the uncertainty faced by households results from 'covariate' shocks, that is, shocks that affect *all* households within a given location, rather than just a few (Alderman & Haque, 2007). In these instances, the returns to back-up livelihood strategies may be minimal because households will diversify into new jobs at the same time. For example, farmers faced with a drought may simultaneously try to supplement their incomes by selling crafts and providing other services. However, during a covariate shock, the demand for these activities is likely to be low, and supply is likely to be high, driving down the returns to diversification.

Also, some households are better able to diversify than others. For example, certain households may lack access to credit for buying the inputs needed to plant new crops or start non-farm businesses. Vitality, it appears that these constraints affect the poorest households disproportionately (Goulden, et al., 2013). This, in itself, does not mean diversification is necessarily a bad indicator of resilience, but it may mean that measures of diversification carry little extra information over and above households' wealth or access to credit.

As such, we can only be certain that diversification constitutes an improvement in resilience if all other factors remain equal. It is, therefore, important to ensure that, in impact evaluations like Oxfam's Effectiveness Reviews, we are able to measure these other factors so that we can consider them in the final analysis.

## 5.5 INDICATORS OF INNOVATION POTENTIAL

We have particularly struggled in the Effectiveness Reviews with finding good indicators of 'Innovation Potential', that is, of whether people have the capability and willingness to experiment and take proactive steps to adapt to change. Early Effectiveness Reviews tried to assess people's attitudes to innovation and change by asking them whether they agreed or disagreed with a series of statements on these subjects. However, these types of questions are subject to the same issues as other perception-based measures (see Section 6) and they tend not to be sensitive enough to detect differences between the intervention and comparison groups.

We have developed two possible alternatives to help capture Innovation Potential in the Effectiveness Reviews. Firstly, in some surveys we have tried asking directly about whether respondents and their households have in fact carried out any new initiatives or innovative activities in the recent past, other than activities that were directly prompted by project activities. We assume that households have engaged in innovations in the past, then they are more likely to have the potential to do so again in the future. Additionally, in more recent surveys, we have tried leaving these questions open, allowing people to describe any initiatives, innovations or changes they have made without trying to provide a list of possible responses in advance.<sup>13</sup> These provide potential avenues for assessing Innovation Potential without relying on perception-based measures per se.

# 6 PERCEPTION-BASED MEASURES

In this section we consider our questions about whether perception-based measures may provide suitable indicators of resilience. We begin by discussing the validity of perception-based measures for *specific* characteristics of resilience. We then look at whether *overall* resilience can be captured using perception-based measures. In both cases, it appears we should exercise caution when using these types of variables in evaluations like the Effectiveness Reviews.

## 6.1 PERCEPTION-BASED MEASURES FOR SPECIFIC CHARACTERISTICS OF RESILIENCE

In the first two years of resilience Effectiveness Reviews, in cases where we were unable to find an objective indicator for a characteristic of resilience, we tended to ask a question about respondents' perceptions of how they scored on that characteristic. For example, we sometimes asked for respondents' perceptions about the fertility of their soil, the quality of pasture land, whether they have any difficulties accessing markets, and the ability of community leaders or local government to respond to crises.

There are, however, at least three potential problems with replacing objective indicators with perception-based measures in this way. For example, when asking a subjective question about soil fertility, it may be that the results reflect a combination of the following:

- The overall sense of optimism/positivity or pessimism/negativity of the respondent.
- The respondents' expectations of what should be considered 'good' soil quality.
- The respondents' perceptions of the purpose of the survey, and whether it is to his or her advantage to give a response that is more positive or negative than the reality.

Of course, this final point potentially affects all the questions in our questionnaires. However, this problem is probably sharper for perception-based measures. For objective measures, it is likely to be more difficult for respondents to 'game' their answers or, at least, there is less reason for us to think that any biases in interview responses to objective questions will be systematically different between the intervention group and the comparison group.

Given these issues, in recent Effectiveness Reviews, we have preferred to omit particular characteristics of resilience from our analysis if they can only be measured subjectively, even if we believe these characteristics to be important.<sup>14</sup> This avoids including a perception-based indicator that may be difficult to interpret and may be correlated only very loosely (and perhaps in a biased way) with the characteristic we would like to measure. Practically, therefore, this has meant excluding soil fertility as a characteristic in the analysis of the 2013/14 Effectiveness Reviews in Pakistan, Mali and Niger, even though this is definitely an important characteristic for livelihoods in each of those areas.

## 6.2 PERCEPTION-BASED MEASURES OF OVERALL RESILIENCE

An increasingly common approach used in evaluation work is to ask people directly about their perception of their own overall resilience. For example, in their document ‘The Resilience Agenda: Measuring Resilience in USAID’ (2013), USAID have directly asked people to describe their household’s ability to ‘cope with and manage future droughts or periods of need/stress’. The main idea behind measures of this type is that people have a good understanding of their capabilities and capacities to deal with disturbance and change, so we do not have to rely on external actors (like impact evaluators) to define what matters for resilience. Additionally, this approach is very direct, so it allows us to reduce the number of indicators required to measure resilience (Jones & Tanner, 2015).

There are three particular concerns with using perception-based measures of overall resilience:

1. Respondents can only answer about their resilience to the extent that they are aware of all the factors affecting their resilience. People’s resilience will be dependent on factors – such as the capacity of local government to support them in times of crisis – about which they may have little knowledge. In a sense, this is a general issue with assessing resilience solely through household-level data. However, the problem is particularly severe if we are asking respondents to define their own overall resilience.
2. It can be difficult to standardise the concepts to which overall resilience questions are referring. For example, we may ask respondents whether they will be able to ‘cope with a future drought’. Clearly, some people may think of a serious drought while others may think of a relatively minor drought. This will depend on the reference event that respondents have in mind when answering the question, which will, in turn, depend on each household’s historical exposure to drought. Equally, respondents may have different ideas of what it means to ‘cope’ with shocks or stresses. This may be particularly problematic if trying to compare project participants with a comparison group, as in the Effectiveness Reviews, because a successful project may lead to people adjusting their standards of what it means to ‘cope’.
3. Finally, as mentioned above, responses to subjective survey questions may depend on respondents’ perceptions of the purpose of the survey. Despite enumerators’ assurances to the contrary, they may believe that the survey is being used to target future support. The potential to distort responses is likely to be larger for subjective measures, and may lead to systematic biases between project participants and the comparison group.

This last point is reinforced by the experience in the first two years of Effectiveness Reviews, where we always asked respondents for their subjective assessment of whether their income had increased or decreased since the (notional) baseline. In most cases, much larger proportions of the intervention group than the comparison group reported experiencing an increase in income, even when there was no evidence of a difference in income based on household consumption, asset ownership, or housing conditions. It appears, therefore, that the subjective question on income change was being used by project participants to express gratitude or satisfaction with the project, rather than providing information about a real income change. A subjective question on resilience could be subject to the same problems.

One possible way of reducing the problems associated with subjective measures of resilience is to use anchoring vignettes. Respondents are given a set of standardised scenarios or households, and then asked to rate the resilience of the fictional households. Then, when the respondents come to rating their own resilience, we have a benchmark against which to compare their answers (Jones & Tanner, 2015; McGregor, et al., 2015; King, et al., 2004).

We have done some work to trial this, and found that using this approach is somewhat challenging. For example, in the baseline and midline surveys that were carried out for the DFID-funded 'Humanitarian Assistance and Resilience-Building in Western Yemen' project, respondents were asked to rate their household's ability to 'respond to shocks and unexpected events', on a scale from one to five. Respondents were also asked to rate the ability of three imaginary households to respond to shocks and unexpected events, using the same scale on which they rated themselves. The three imaginary households were presented in the questionnaire using standardised descriptions in order to provide a set of reference points. Respondents' ratings of the resilience of the households in each of those three scenarios were then compared with their rating of their own household's level of resilience, in order to create a score for household resilience that is comparable between respondents and over time. However, even with this calibration procedure, it was difficult to create an indicator of resilience that was sensitive enough to show any meaningful changes between groups.

## 7 CONCLUSIONS AND NEXT STEPS

In this discussion paper, we have documented some of the main lessons learned from four years of conducting Effectiveness Reviews on projects aimed at building resilience. We first described how focusing on drivers or characteristics enables us to measure resilience with a snapshot, without relying on shocks or stresses actually occurring, and considered how best to disentangle this causal chain. We then examined the implications of Oxfam's multidimensional approach, focusing in particular on how to set thresholds and aggregate resilience indicators. Finally, we turned to some practical issues, explaining the most successful methods for actually generating indicators given the conceptual framework, and highlighting some of the potential perils of using subjective measures for the Effectiveness Reviews.

How useful will these lessons be in future Effectiveness Reviews, and how relevant will they be for others looking to develop strategies to evaluate resilience-building projects? Many of the practical issues around developing indicators, be it finding tools to structure discussions or the difficulties with subjective indicators when conducting *ex post* evaluations with comparison groups, should be applicable to many other settings. Moreover, the difficulties associated with measuring resilience using a single snapshot after a project has reached maturity, but not necessarily after shocks or stresses have occurred, should be familiar to other practitioners working in monitoring and evaluation.

Oxfam is currently working to try to address some of the issues outlined in this discussion paper, especially around the aggregation step required to build the resilience index. In particular, we have begun working on new methods for constructing weights for each indicator, by trying to elicit respondents' preferences over each indicator or each dimension. This involves both participatory approaches, in which the relative importance of indicators or dimensions are discussed in focus groups, as well as augmenting our surveys with questions which ask respondents to express their preferences over indicators or dimensions directly. We hope that these methods will result in an index that better reflects what resilience means to the participants in Oxfam projects.

Alongside our work to strengthen our evaluation design, there have also been efforts to refine how we define 'resilience' and support effective programming. More recent guidance for the design and implementation of Oxfam projects has conceptualised resilience in terms of three capacities, namely, adaptive, absorptive and transformative, as set out in Box 1. (Jeans, et al., 2015). This approach has become common not just within Oxfam, but across the sector as a



whole. As we move forward, we will therefore also be looking at the implications of moving from five dimensions to three capacities.

### Box 1: Absorptive, Adaptive and Transformative Capacity in Oxfam Programming

- 1. Absorptive Capacity** – the capacity to take protective action (including preparedness) and to cope with shocks and stress. It is needed as shocks and stress will continue to happen, for example due to climate variability, protracted conflict and extreme weather events.
- 2. Adaptive Capacity** – the capacity to make adjustments and incremental changes in anticipation of or in response to change, in ways that create more flexibility in the future. It is needed as change is ongoing and uncertain, and because systemic change takes time and sustained engagement.
- 3. Transformative Capacity** – the capacity to change systems that create risk, vulnerability and inequality. It is needed to address the drivers of risk, vulnerability and inequality and because social and natural systems are themselves being transformed, for example by globalisation and climate change.

Much of the work in the Effectiveness Reviews will be suitable for this new framework. Previous indicators for both drivers/characteristics of resilience and of final outcomes (assuming shocks or stresses have already occurred), should capture absorptive capacity, providing the lessons outlined in this discussion paper can be put in practice. Also, in trying to measure Innovation Potential, we have already begun to tackle some of the problems associated with finding indicators for adaptive capacity.

However, fresh challenges arise with trying to measure transformative capacity. Firstly, transformation generally takes far longer than both absorption and adaptation. This longer time horizon may reduce our confidence in the causal links from drivers or characteristics of resilience to final outcomes, that is, realisation of rights and improvements in wellbeing. Without evidence to support these causal relationships, we cannot be sure our supposed indicators of transformative capacity are valid drivers or characteristics of resilience. Moreover, as we mentioned in Section 3.1, in some Effectiveness Reviews it has been possible to measure realisation of rights or wellbeing shortly after a shock or stress had occurred, to provide extra indicators of resilience. However, measuring ‘applied capacity’ in this way is not likely to be possible for transformative capacity, since the relevant stresses are long term.

Transformative capacity also requires measurement at a different scale from absorptive and adaptive capacity, given its focus on system-level change. Household-level data may give some indication about the extent to which communities, regions or eco-systems are resilient, but this type of large-N data will necessarily focus on the unit of the individual or household. In this way, indicators of this kind may better capture the ways in which transformation assists absorption or adaptation, rather than assessing transformative capacity directly. For example, we may ask households whether or not they have access to a formal government safety net in times of crisis to try and measure the resilience of the political system. However, rather than capturing transformative capacity per se, this may instead pick up the households’ ability to absorb shocks and stresses. To understand the resilience of the political system, we would benefit from collecting small-N data *at the level* of the system, rather than gathering data from that system’s constituent parts. We are aiming to use this approach, using the methodology of ‘process tracing’ in particular, for the 2015/16 wave of Effectiveness Reviews.<sup>15</sup>

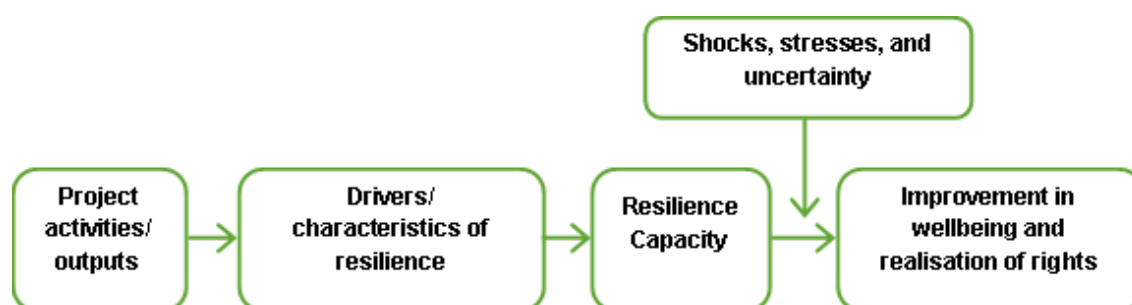
The next steps for the resilience Effectiveness Reviews, therefore, should ensure we can generate indicators for the three capacities, especially transformative. We recognise that the large-N quantitative indicators on which we have relied thus far may be complemented by using small-N qualitative data at the level of the system. This presents an exciting avenue for future research into measuring resilience in *ex post* evaluations.

# NOTES

<sup>1</sup> The number of household surveys conducted ranges from 400–800 households.

<sup>2</sup> In the Oxfam International Resilience Metrics subgroup we have used a slightly different diagram, with overall resilience capacity being distinguished from the drivers of resilience (see Figure 3). It may be that resilience capacity can be measured more directly through self-perception measures, in spite of the caveats outlined in Section 6.

**Figure 3: An Alternative Timeline for Measuring Resilience**



<sup>3</sup> For example from the Effectiveness Review series, see the reports from Kenya and Ethiopia.

<sup>4</sup> See, for example, the Effectiveness Reviews evaluating projects aiming to build resilience against flooding in Nepal and Pakistan.

<sup>5</sup> See, for example, the Effectiveness Reviews in Mali, Niger and Chad.

<sup>6</sup> We consider participatory approaches to assigning weights in Section 5.2.

<sup>7</sup> This was the aggregation method used by earlier resilience Effectiveness Reviews.

<sup>8</sup> See, for example, the 2012/13 Effectiveness Review on Ethiopia.

<sup>9</sup> Put differently, we have tended to do one-stage weighting rather than two-stage weighting in the most recent Effectiveness Reviews.

<sup>10</sup> For examples, see the Effectiveness Reviews in Chad and Niger. The Venton (2013) tool was also used in developing the proposal for BRACED funding in Myanmar in July 2014, where one of the partner organisations, Spectrum, conducted focus groups based on these questions in four communities in two different areas of the country.

<sup>11</sup> This has happened both in Myanmar and in Chad.

<sup>12</sup> Other questions included in the Venton tool we have found useful as follow-ups are:

- Are some factors cited more than men or more by women?
- How does the typical household in the community rank on these factors?
- What should the community do to be more resilient?

<sup>13</sup> See, for example, the Effectiveness Reviews in Nicaragua and Chad.

<sup>14</sup> This is consistent with the question raised by Levine (2014) that 'If dimensions which are hard to pin down ... are important, what justification is given for believing that, just because it would be useful if they could be measured, there is any validity in indicators that purport to do this?'

<sup>15</sup> For more information on Oxfam's approach to process tracing, see <http://policy-practice.oxfam.org.uk/blog/2013/02/~media/C396B507E01C47AB880D7EEF9ECCD171.ashx>.

# BIBLIOGRAPHY

- Alderman, H. & Haque, T., 2007. Insurance Against Covariate Shocks: The Role of Index-Based Insurance in Social Protection in Low-Income Countries of Africa. *World Bank Working Papers*, Volume 95.
- Beegle, K., Dehejia, R. H. & Gatti, R., 2006. Child Labor and Agricultural Shocks. *Journal of Development Economics*, Volume 81, pp. 80–96.
- Béné, C., 2013. Towards a Quantifiable Measure of Resilience. *IDS Working Paper*, Issue 434.
- Béné, C., Frankenberger, T. & Nelson, S., 2015. Design, Monitoring and Evaluation of Resilience Interventions: Conceptual and Empirical Considerations. *IDS Working Paper*, Issue 459.
- Bushell, H. & Hughes, K., 2013. *A Multidimensional Approach for Measuring Resilience*, s.l.: Oxfam.
- Dercon, S., 2008. Fate and Fear: Risk and Its Consequences in Africa. *Journal of African Economies*, Volume 17, pp. 97–127.
- Frankenberger, T. R., Constan, M. A., Nelson, S. & Starr, L., 2014. How NGOs Approach Resilience Programming. In: S. Fan, R. Pandya-Lorch & S. Yosef, eds. *Resilience for Food and Nutrition Security*. Washington, DC: International Food Policy Research Institute, pp. 177–186.
- Goulden, C. G., Adger, W. N., Allison, E. H. & Conway, D., 2013. Limits to Resilience from Livelihood Diversification and Social Capital in Lake Social-Ecological Systems. *Annals of the Association of American Geographers*, 103(4), pp. 906–024.
- Hiller, D. & Castillo, G. E., 2015. *No Accident: Resilience and the Inequality of Risk*, Oxford: Oxfam.
- Holland, J., 2013. *Who Counts? The Power of Participatory Statistics*. Rugby, UK: Practical Action Publishing.
- Hutchings, C., 2014. Balancing accountability and learning: a review of Oxfam GB's global performance framework. *Journal of Development Effectiveness*, 6(4), pp. 425–435.
- Jeans, H., Thomas, S. & Castillo, G., 2015. *Oxfam Framework and Guidance for Resilience Programming – Draft*, Oxfam.
- Jones, L. & Tanner, T., 2015. Measuring 'Subjective Resilience': Using People's Perceptions to Quantify Household Resilience. *ODI Working Paper*, Issue 423.
- King, G., Murray, C., Salomon, J. & Tandon, A., 2004. Enhancing the Validity of Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review*, 98(1).
- Levine, S., 2014. Assessing Resilience: Why Quantification Misses the Point. *Humanitarian Policy Group Working Paper*.
- Magnan, A., 2013. Éviter la maladaptation au changement climatique. *IDDRI Policy Briefs*, Volume 8.
- McGregor, A., Coulthard, S. & Camfield, L., 2015. Measuring what matters: The role of well-being methods in development policy and practice. *Overseas Development Institute Project Note*, Issue 04.

Schipper, E. L. F. & Langston, L., 2015. A Comparative Overview of Resilience Measurement Frameworks: Analysing Indicators and Approaches. *Overseas Development Institute Working Paper*, Issue 422.

USAID, 2013. *The Resilience Agenda: Measuring Resilience in USAID*, Washington, DC.

Venton, P., 2013. *Analyse Paul Venton – Outil de Collecte*.

## Oxfam Discussion Papers

Oxfam Discussion Papers are written to contribute to public debate and to invite feedback on development and humanitarian policy issues. They are 'work in progress' documents, and do not necessarily constitute final publications or reflect Oxfam policy positions. The views and recommendations expressed are those of the author and not necessarily those of Oxfam.

This paper was written by Rob Fuller and Jonathan Lain. For more information, or to comment on this report, email [ppat@oxfam.org.uk](mailto:ppat@oxfam.org.uk)

© Oxfam GB December 2015

This publication is copyright but the text may be used free of charge for the purposes of advocacy, campaigning, education, and research, provided that the source is acknowledged in full. The copyright holder requests that all such use be registered with them for impact assessment purposes. For copying in any other circumstances, or for re-use in other publications, or for translation or adaptation, permission must be secured and a fee may be charged. E-mail [policyandpractice@oxfam.org.uk](mailto:policyandpractice@oxfam.org.uk).

The information in this publication is correct at the time of going to press.

Oxfam GB, Oxfam House, John Smith Drive, Cowley, Oxford, OX4 2JY, UK.

# OXFAM

Oxfam is an international confederation of 18 organisations networked together in over 90 countries, as part of a global movement for change, to build a future free from the injustice of poverty:

Oxfam America ([www.oxfamamerica.org](http://www.oxfamamerica.org))  
Oxfam Australia ([www.oxfam.org.au](http://www.oxfam.org.au))  
Oxfam-in-Belgium ([www.oxfamsol.be](http://www.oxfamsol.be))  
Oxfam Brasil ([www.oxfam.org.br](http://www.oxfam.org.br))  
Oxfam Canada ([www.oxfam.ca](http://www.oxfam.ca))  
Oxfam France ([www.oxfamfrance.org](http://www.oxfamfrance.org))  
Oxfam Germany ([www.oxfam.de](http://www.oxfam.de))  
Oxfam GB ([www.oxfam.org.uk](http://www.oxfam.org.uk))  
Oxfam Hong Kong ([www.oxfam.org.hk](http://www.oxfam.org.hk))  
Oxfam India ([www.oxfamindia.org](http://www.oxfamindia.org))  
Oxfam Intermón (Spain) ([www.oxfamintermon.org](http://www.oxfamintermon.org))  
Oxfam Ireland ([www.oxfamireland.org](http://www.oxfamireland.org))  
Oxfam Italy ([www.oxfamitalia.org](http://www.oxfamitalia.org))  
Oxfam Japan ([www.oxfam.jp](http://www.oxfam.jp))  
Oxfam Mexico ([www.oxfammexico.org](http://www.oxfammexico.org))  
Oxfam New Zealand ([www.oxfam.org.nz](http://www.oxfam.org.nz))  
Oxfam Novib ([www.oxfamnovib.nl](http://www.oxfamnovib.nl))  
Oxfam Québec ([www.oxfam.qc.ca](http://www.oxfam.qc.ca))

Please write to any of the agencies for further information, or visit [www.oxfam.org](http://www.oxfam.org).

