Measuring Resilience Impact

at Programme and Project Levels

July 2016

christian aid

Author:

Richard Ewbank, Climate Adviser

Acknowledgements:

Thanks to Nicholas Abuya, Richard Ashiagbor, Kate Bingley, Gerhard Buttner, James Cannon, David Clark, Peter Crawford, Jane East, Louise Hart, Amy Ingham, Sabine Joukes, Ram Kishan, Alicia Malouf, Sophie Makoloma, Josephine Matriano, Michael Mosselmans, Eston Njuki, Emily Schechter, Fran Scott, Allan Vera and Simone di Vincenz for their expert advice and feedback.

Christian Aid is a Christian organisation that insists the world can and must be swiftly changed to one where everyone can live a full life, free from poverty.

We work globally for profound change that eradicates the causes of poverty, striving to achieve equality, dignity and freedom for all, regardless of faith or nationality. We are part of a wider movement for social justice.

We provide urgent, practical and effective assistance where need is great, tackling the effects of poverty as well as its root causes.

christianaid.org.uk

Contact us

Christian Aid
35 Lower Marsh
Waterloo
London
SE1 7RL
T: +44 (0) 20 7620 4444
E: info@christian-aid.org
W: christianaid.org.uk

Contents

Cover: A community in Bihar, India, use participatory community mapping to illustrate the risks they face and the resilience they have built as part of an intervention supporting the recovery of their livelihoods after the catastrophic Kosi river floods in 2008.

Photographs: Richard Ewbank and David Clark (Kenya)

Foreword

This “how to” guide draws on many years of programming experience of Christian Aid and our partner organisations, to provide an important contribution to the wider sector as we grapple with the challenge of how best to measure the impact of resilience programming.

Christian Aid’s approach highlights the importance of programme participants defining resilience, and related indicators, in their own terms, as they understand it in their own context. In making the link between regular monitoring and deeper explorations of impact, we also recognise the empowerment potential that comes with promoting processes of self-evaluation.

This guide offers a wealth of information, from considerations around data collection and analysis through to communication and utilisation of findings. I strongly encourage practitioners to engage with this resource.

Kate Bingley
Co-Head of Research, Evidence and Learning

Christian Aid
July 2016

Below: As part of a resilience-building process, women in Burkina Faso identify and prioritise the risks that most seriously affect their lives and livelihoods

1. Introduction

Although any monitoring and evaluation (M&E) system tends to focus on outputs, outcomes and even 'value for money', it is important to emphasise that from a participant[1] and programming perspective, it is **impact** that determines the eventual usefulness of an intervention.

Impact has been defined quite loosely– *'improvements in the lives and livelihoods of beneficiaries'*– as well as more specifically – *'impact concerns long-term and sustainable changes introduced by a given intervention in the lives of beneficiaries. Impact can be related either to the specific objectives of an intervention or to unanticipated changes caused by an intervention; such unanticipated changes may also occur in the lives of people not belonging to the beneficiary group. Impact can be either positive or negative, the latter being equally important to be aware of.'* [2]

This second definition usefully highlights issues of expected and unexpected, as well as both positive and negative impact. In logical framework terms, impact has more recently been described as the result of outcomes that are, in turn, the direct result of outputs, i.e. 'outcomes' have been inserted between outputs and impact. Some practitioners prefer to describe outcomes as impact, or 'immediate impact'. It is recognised that the definition of 'outcome' and 'impact' can have some overlap and the boundary is a bit fuzzy.

Impact is more likely to be achieved, measured and documented if a project implements a **managing for impact** approach. This implies:

- *establishing information-gathering and management mechanisms* – ensuring that systems are in place to provide the information needed from the outset, encourage learning and facilitate project impact-enhancing management and actions going forward
- *creating an environment for learning* – establishing relationships with all involved in an inclusive way (ensuring equal weight is given to the women and men involved) in order to build trust, stimulate innovation and foster commitment
- *guiding the strategy/theory of change* – assessing whether an initiative is consistent with its theory of change, so providing evidence on the validity of the assumptions made about how inputs generate outputs that deliver outcomes that result in impact. If this relationship is not working, adjustment in the strategy/theory of change and even the overall objectives will be needed
- *supporting dialogue and advocacy* – delivering credible evidence on impact that promotes scale-up and enlightens other stakeholders, especially those involved in relevant local, national and international policy and decision-making.

'The counter-bureaucracy ignores a central principle of development theory – that those development programs that are most precisely and easily measured are the least transformational, and those programs that are most transformational are the least measurable'

Andrew Natsios (USAID Administrator, 2001-06)

Figure 1: Managing for impact [3]

Elements	Organisation that manages for impact	Organisation that does not manage for impact
A focus on strategy and theory of change – understanding the goals and objectives of the organisation and its projects, allocating available resources and guiding relations between stakeholders to maximise impact.	The implementing partners and primary stakeholders collect impact data and meet regularly to consider if impacts can be seen and whether they meet expectations.	The organisation leaves impact assessment to outside organisations and only focuses on monitoring the implementation of activities.
A learning environment – inspiring and helping those involved with the project to reflect critically on progress, to learn from mistakes and to generate ideas for making improvements.	Projects and partner staff and primary stakeholders regularly and openly discuss progress and problems. Problems and mistakes are the basis for learning how to work better. People feel safe enough to share their ideas and observations with higher authorities and peers.	Staff focus only on implementing activities without asking if they are relevant to local poverty needs. Project and partner staff feel that their ideas and observations are not valued and do not speak up when they see things going wrong. Primary stakeholders are rarely asked for their opinions.
Effective operations – planning, organising and checking staff inputs, equipment, partner contracts, financial resources, workplans and communications to implement activities effectively and efficiently.	Clear systems exist for tracking staff performance, equipment use, etc. People are clear about their responsibilities and deadlines and the resources they need to do the work are available. Managers can access/allocate sufficient budgets and are building capacity where needed.	People are often unclear about what they should be doing, when and where. Vehicles and other resources are poorly coordinated and often not available on time. Much time is wasted in inactivity. No efforts are made to fill capacity gaps.
A supportive M&E system – designing and implementing information gathering and reflective learning processes to generate insights that help improve operations and strategic directions.	Managers have quick and easy access to information about project progress, including budget monitoring data. Staff and partners generate information about project outputs and impact, and this information is regularly shared with all stakeholders in an appropriate way. The walls of project offices show up-to-date information about progress, including graphs, charts, photographs, etc. Adhoc studies are undertaken to explain any emerging problems. Project reports are interesting, telling both the good and bad and how improvements will be made.	Few people know what the project has achieved to date. There is little evidence about whether all the activities of the project are leading anywhere. M&E is seen largely as an external reporting function, e.g., to satisfy a donor rather than as an important management tool. Project reports are uninteresting (and therefore a burden to compile), are not analytical, exaggerate success and do not mention problems. Little information is shared with project stakeholders.

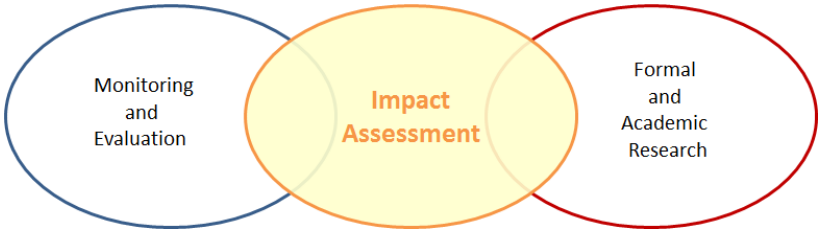
Below: Community leaders in Mali explain to evaluators the impact of their activities to improve the sustainable management of local forests and how this enhances their resilience to climate change. Finding a safe, appropriate and comfortable location where all participants feel empowered to contribute fully is the key to a relaxed and open conversation.

Although the advice in this guideline tends to fit the typical impact assessment carried out by an external assessor (such as a project evaluation officer or a consultant), the basic principles and tools equally apply to empowering a community-based organisation, e.g., a village development committee to measure their own resilience. In fact, although this form of assessment tends not to get significant attention, it is the more important in promoting resilience-building processes.

It is also important to emphasise from the outset that while impact often takes time to emerge, making impact assessment more important later rather than earlier in the project cycle, certain types of **resilience impact may emerge relatively quickly.** So farmers using a weekly forecast through the first season of a project may benefit from impact as they make informed decisions – this makes an assessment at the end of each season an important information gathering opportunity. The nature of this resilience building as a process suggests that forms of continuous data collection on key measures of resilience, for example at monthly or quarterly meetings with participants, are more likely to capture impact than occasional assessments at the mid-term or end of an intervention.

Impact assessment is distinct from M&E and formal research, but draws from both, acting as the bridge between them:

Figure 2: The link between M&E and research



This guidance focuses primarily on impact assessment of projects and programmes designed to increase the resilience of vulnerable people and their communities. Resilience of other assets, such as landscapes, ecosystems and infrastructure, is no less important but in-depth analysis of their resilience does present more specific technical challenges that are beyond the central focus of this document. Guiding the planning and design of general M&E systems that may include impact assessment is also not the purpose of this resource, but the reference link (see Annex 1) does include a folder (general guidelines) with material that addresses this issue. (It should also be noted that this is a guideline not a tramline.)

Annex 1 highlights reference sources both resilience-specific and general with an intranet link to these and the tools listed in annex 2.

Annex 2 briefly describes each of these tools.

2. Resilience concepts in measuring impact

In Partnership for Change,[4] resilience was defined as 'the power of individuals and communities to live with dignity, responding successfully to disasters and the opportunities and risks they face'. It is a capacity-building process to enhance the ability of individuals and communities to 'anticipate, organise for and adapt to change'.

Folke at al. state that 'resilience, for social-ecological systems, is related to (a) the magnitude of shock that the system can absorb and remain within a given state, (b) the degree to which the system is capable of self-organization, and (c) the degree to which the system can build capacity for learning and adaptation. Management can destroy or build resilience, depending on how the social-ecological system organizes itself in response to management actions.' [5]

Drawing on these principles, from a humanitarian and development perspective, five characteristics – adaptive capacity, buffers, robustness, recovery and thriving (as per Figure 3 below) – are the key factors in what makes an individual, community or livelihood resilient. Assessing the impact of any intervention that seeks to increase the resilience of an organisation, a community or group of individuals would therefore involve measuring change in one or more of these.

2.1 Buffer capacity

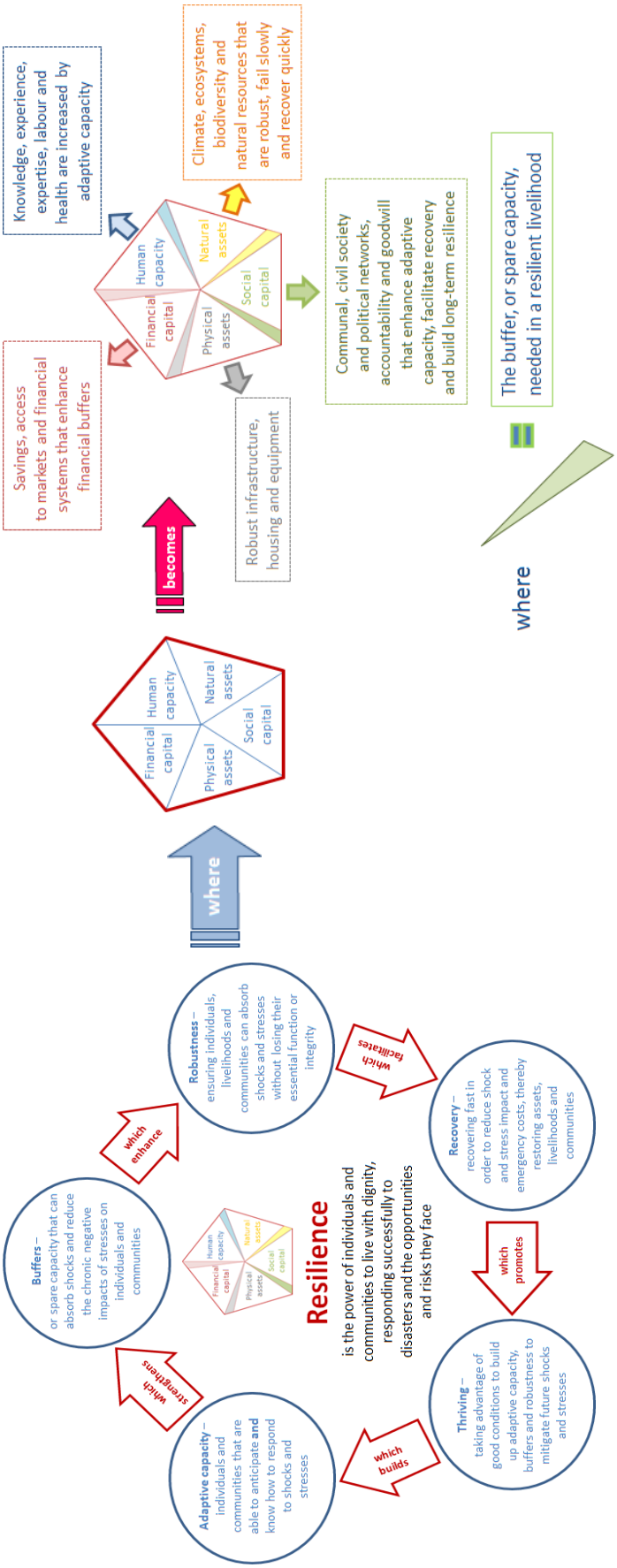
Adaptive capacity, robustness, recovery and thriving all depend on the assets available – the need to develop, strengthen or maintain the ability of assets to mitigate the effects of shocks and stresses puts a particular emphasis on the ability of an intervention to strengthen asset redundancy or buffer capacity, thereby developing a *resilient asset pentagram* (as per Figure 3 below). This buffer capacity is what enables a livelihood to cope in times of shock and is more gradually built to address the future consequences of incremental stress.

So, when understanding the assets that people or communities have, a resilient livelihood relies on a balanced and mutually reinforcing mix that incorporates an adequate level of buffer capacity, illustrated by the additional slices added to each asset category. Higher levels of resilience are achieved when the *optimal level of redundancy* (or buffers) is reached. This level changes according to circumstances, demonstrating the importance of seeing resilience building as a continuous process rather than a linear path to a one-off objective. It follows therefore that any assessment of the impact of resilience-building activities that seeks to assess whether or not progress is being made towards this optimal level should also be continuous,[6] providing regular information that can modify objectives, improve resilience building and minimise the possibility that an intervention is unintentionally making people more vulnerable (through maladaptation).

'It is not the strongest who survive, nor the most intelligent, but the ones most adaptable to change'

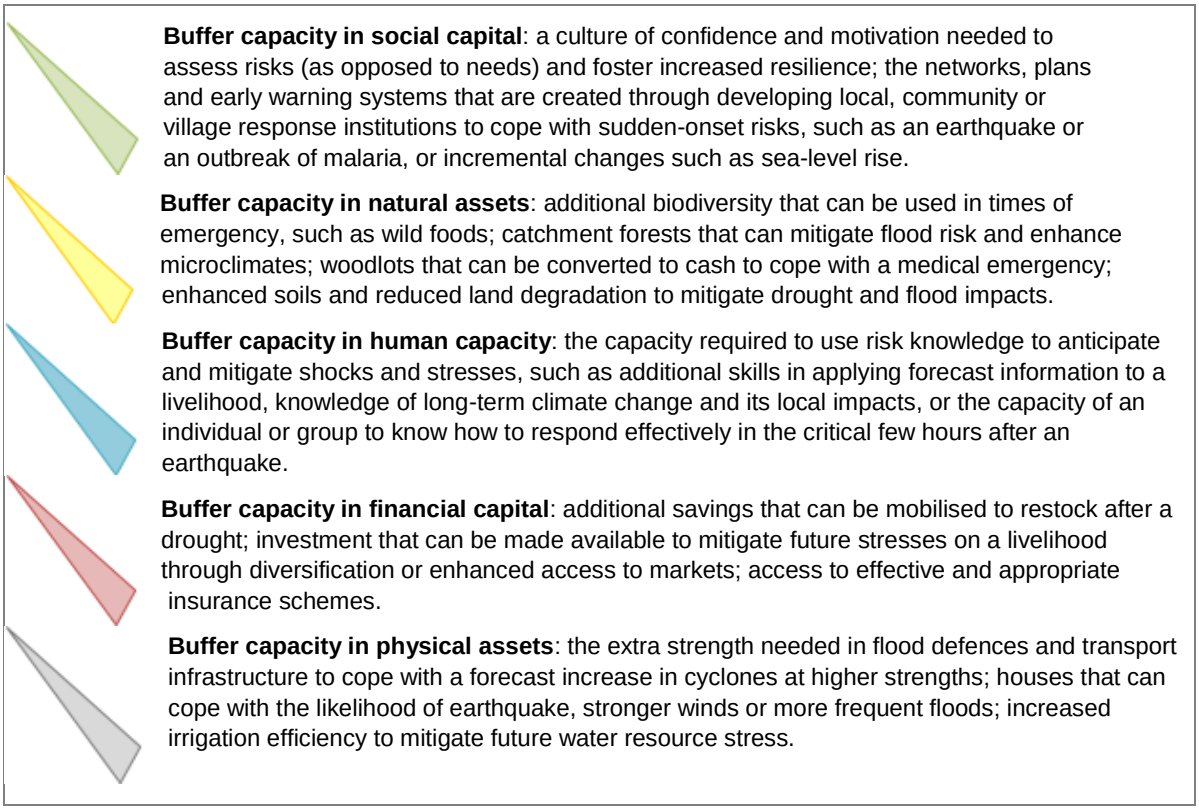
Charles Darwin

Figure 3: The components of resilience and resilient assets [7]



Examples for each asset category include:

Figure 4: Asset buffers



2.2 Standby and parallel buffers

While the resilient asset pentagram illustrates the difference between 'normal' or everyday assets and buffers to emphasise the importance of redundancy in resilience, in practice buffers also are differentiated according to whether they are only used when a shock emerges or are in at least partial use during average as well as unusual conditions. So here the distinction is between:

- parallel buffers – where buffer capacity exists in a livelihood that can be used in response to shocks and stresses, but is in at least partial use during normal conditions
- standby buffers – a spare unit, resource or procedure that is **only** used when a shock is anticipated or emerges.

The boundaries between parallel buffers and everyday assets will not be rigid. The extent to which a livelihood can effectively integrate everyday assets, parallel buffers and standby buffers will therefore have a significant impact on its resilience. So medical skills developed as part of building capacity to respond to earthquakes may well be equally relevant to other smaller-scale shocks and even applied to everyday health wellbeing. Alternatively, the two metres added to a flood protection bank to cope with the extra degree of flooding anticipated through climate change will only prove a useful asset when those heavier floods occur, i.e. it is a standby buffer.

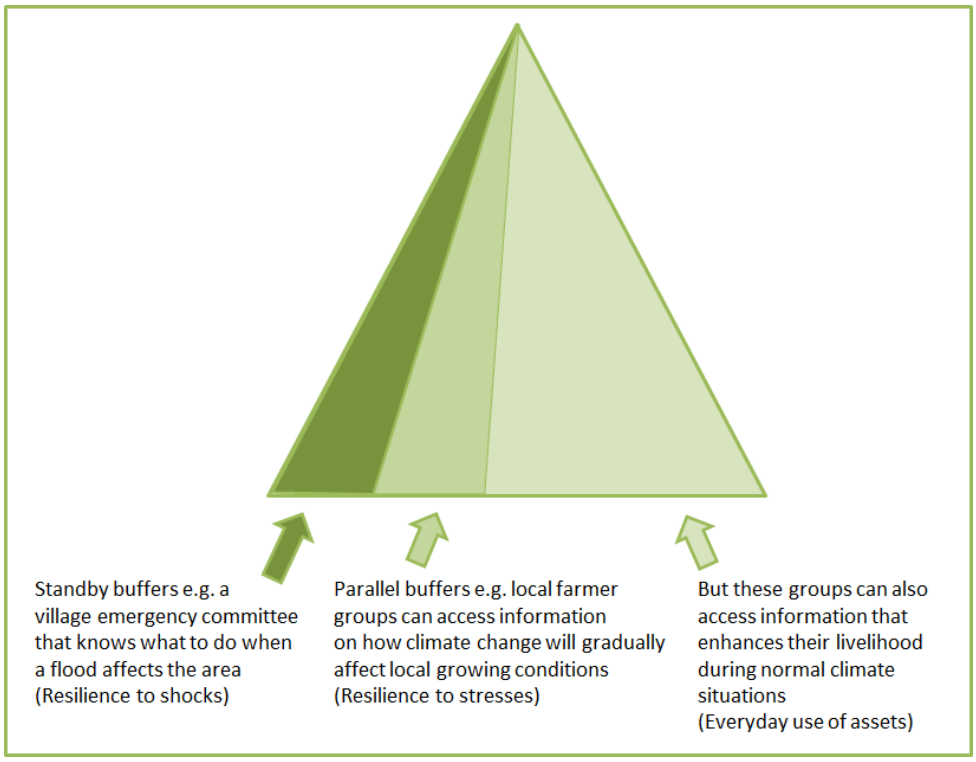
Measuring the impact of this extra protection will have to wait until the flood happens, which has important implications for the planning and implementation of impact assessment activities. If projects adhere to a rigid M&E workplan and do not implement adhoc post-

shock impact assessments, they are likely to leave a significant aspect of resilience impact poorly measured or even unmeasured.

In general, understanding the impact of resilience building with respect to standby buffer capacity is likely to be more important when assessing resilience to specific shock and hazards; on the other hand, understanding parallel buffer capacity is likely to be as important when assessing resilience to incremental stresses and to shocks. For example, when considering social capital, the standby buffer of an emergency committee that has been trained in flood response will be an asset that, like the extra two metres on the flood bank, remains unused until a flood occurs. On the other hand, a network of community groups, such as farmer groups that have better access to meteorological information, is likely to both increase resilience to the shocks and stresses of climate change and enhance crop performance in normal or good conditions. The network of farmer groups is as much in use for everyday livelihood management as it is for greater climate change resilience (see Figure 5 below).

Understanding how resilience impact is achieved in, and differs between, parallel and standby buffers can assist the development of useful recommendations for further resilience-building. For example, persisting with strengthening standby buffers has implications if there are long intervals between shocks that cause standby buffer capacity to degrade over time. If this degradation is underestimated, a false sense of security can reduce rather than increase resilience, as shown by the failure of the cyclone protection infrastructure in New Orleans during Hurricane Katrina.

Figure 5: Standby and parallel buffers in social capital



The same kind of analysis can also be used to assess the other four asset categories, for example:

Figure 6: Standby and parallel buffers in other assets

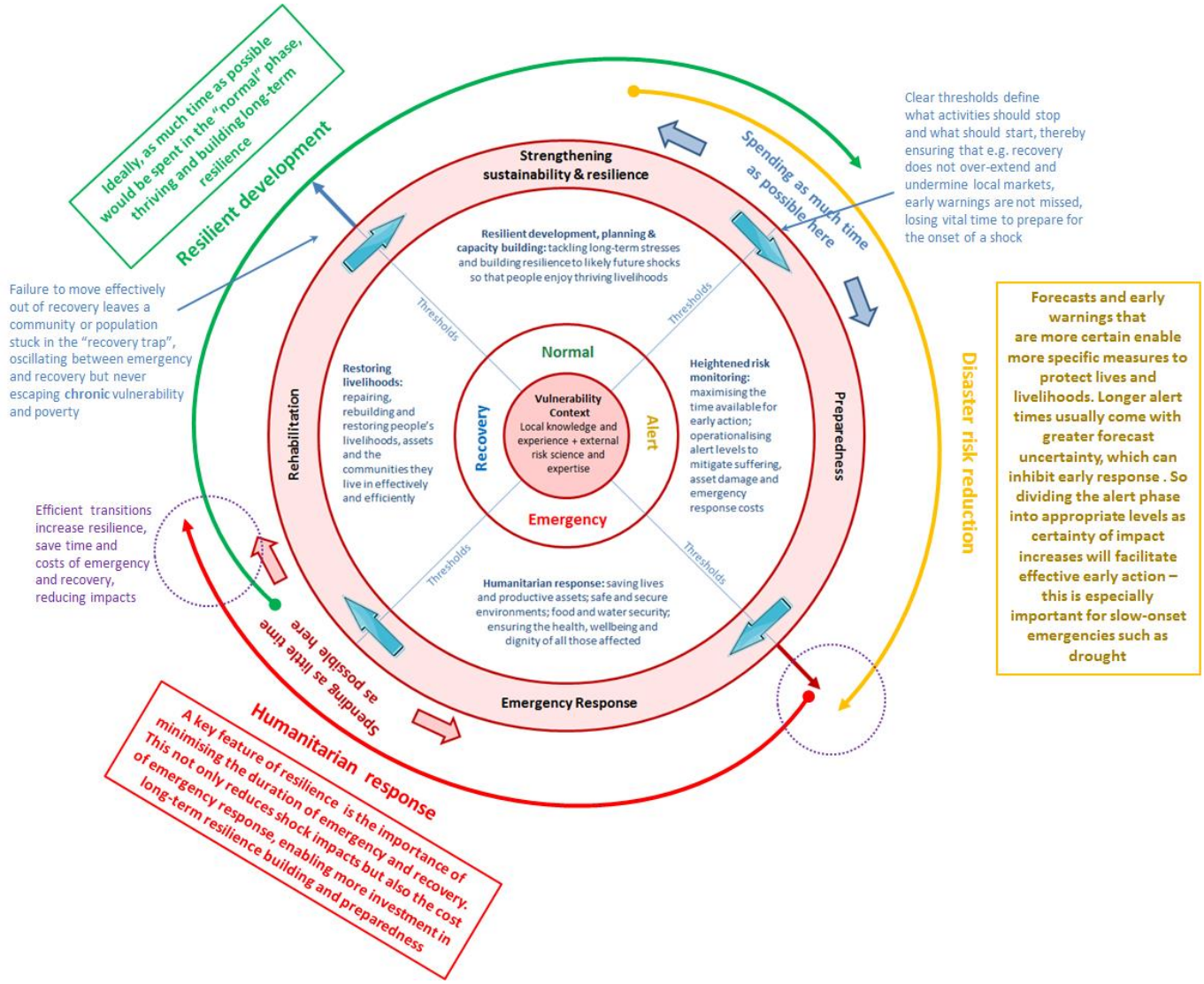
Type of asset	Everyday assets	Parallel buffers	Standby buffers
Natural	Soils that are in use for food production and security.	Catchment forests that regulate the local microclimate, as well as mitigate flood and landslide risks in response to heavy rainfall or provide emergency food items in time of drought.[8]	Biodiversity that is conserved in case it contains valuable genetic diversity for future adaptation to changing weather and disease patterns (crops, medicines, etc.).[9]
Human	Understanding of normal local weather patterns that support relevant livelihoods.	Knowing how to use weather forecasts that guide livelihood decisions in relation to rainfall, as well as give an understanding of when flood risks are rising.	Knowledge of how to respond to a flood early warning that ensures lives and assets are saved and protected.
Financial	Finances under the mattress or in a current account.	Savings accounts that are dipped into on special occasions, but also in times of emergency.	Purchase of an insurance policy that only pays out after a shock occurs.
Physical	Irrigation channels that are used annually for enhanced crop production.	Increased efficiency that reduces cost in normal years, but also ensures that enough water is available during dry spells.	The extra capacity required to ensure that enough irrigation water can be stored for a severe drought year.

Buffers can present a challenge to conventional notions of economic efficiency, which call for maximising the productive use of all assets.[10] To this extent, uncritically enhancing marketisation of assets in the short-term can undermine longer-term resilience by eroding buffers or preventing their development. Although businesses large and small across the world are recognising the increased costs of climate change, few are investing in the value-chain resilience-building measures needed.[11]

Relating this to the risk management cycle (see Figure 7 below) shows how an understanding of the five key resilience factors guides an assessment of what to measure to understand if resilience has been enhanced, such as through a participatory vulnerability and capacity assessment (PVCA) process. A PVCA should take into account assets and especially the buffers that are already in use – an impact assessment approach would then explicitly seek to measure how these buffers are being strengthened given the risk scenario developed, how they need to be enhanced in 'normal' conditions to maximise resilience to shocks and stresses, and how they facilitate efficient transition through the risk cycle when shocks occur.

The risk cycle also demonstrates the importance of adaptive capacity – anticipating shocks and stresses and knowing how to respond to anticipation – to building resilience. In much development practice, the importance of early action during the alert phase is underestimated.

Figure 7: Risk management cycle



A shock event should be seen as an important opportunity to measure resilience, ideally later in the emergency response or during the recovery phase. Even without this happening, resilience to shocks can be 'stress tested' against likely shock scenarios, for example through flood drills.[12] Ideally both should be carried out in any intervention that seeks to increase resilience to shocks.

In the cycle, increasing resilience implies reducing the amount of time a livelihood spends in the bottom left-hand half of the cycle (emergency/recovery) and maximising that spent in the top right-hand side (normal/alert), so increased lead times before a shock occurs, reduced cost and time spent in emergency, and increased speed of recovery after will all be important characteristics. This will increase the time and resources available in 'normal' circumstances to build resilience to longer-term incremental stress and ensure that travelling around the risk cycle with the next shock is more efficient and less traumatic.

2.3 Importance of costs saved and losses avoided

Another challenge in measuring the impact of resilience is that increased resilience does not necessarily mean increased assets at all times – for example, a livelihood with higher levels of assets but no buffers may be less resilient than one with less assets but better-developed buffer resources. Likewise a more resilient livelihood may go backwards at certain periods, such as after a significant shock, but just not as much as a non-resilient livelihood with less well-developed buffers and robustness. The key issue here is how effective buffers are in absorbing the shock and enabling rapid recovery. With a more incremental stress, the focus will be more on how parallel buffers enable a similarly incremental process of adaptation in order to mitigate the stress.

This is explicitly recognised when applying measures of cost effectiveness to disaster risk reduction. These measures typically focus on the costs saved from anticipatory humanitarian intervention – so livelihoods affected by a shock still require external support, but this is reduced for those that are well prepared in advance, incorporate more long-term resilience and are able to take maximum advantage of the alert phase. As this shows, the nature of resilience implies as much a focus on damage avoided and costs saved as an impact, as on benefits gained (or thriving). The challenge for damage avoided is in measuring something that did not happen as opposed to something that did.

Improving resilience is not simply a process of relentless growth or positive impact, as assumed by most logic-based linear planning and assessment processes. Consequently measuring resilience is not a simple process of measuring assets and assuming that if they are increasing, so is resilience, but measuring assets and their capacity to buffer lives and livelihoods against shocks and stresses.

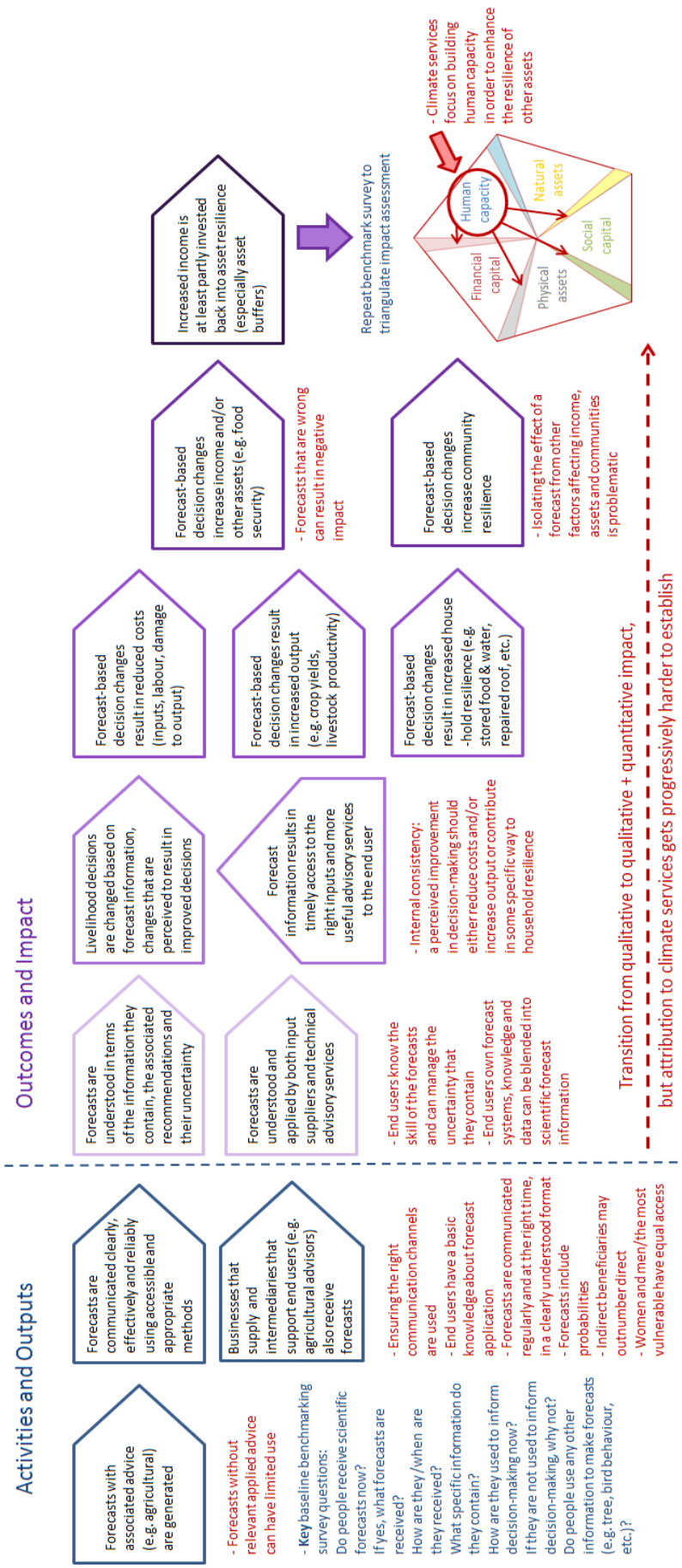
Measuring resilience is not a simple process of measuring assets and assuming that if they are increasing, so is resilience, but measuring assets ***and their capacity to buffer lives and livelihoods against shocks and stresses***

2.4 Impact pathways

Impact generally follows a pathway of cause and effect,[13] with outcome as the first level of impact (or immediate impact). For example, training on use of climate forecasts would aim to have as a key outcome an increased understanding by project participants of the forecast. But this is unlikely to be the last word in terms of training impact. An impact assessor could reasonably ask *'so what?'* and look for a second level of impact. In this case (see Figure 8 below), increased access to and understanding of a forecast could result in a second level of impact – that decisions have been changed as a result.

But still, the concern is *'so what – how has this changed decision resulted in increased resilience?'* A changed decision could result in actions that increase productivity, decrease the use of inputs (thereby increasing efficiency), or avoid damage. So some clear (and in this case, quantitative) measures are starting to emerge. But even here (at the third level), the story does not end.

Figure 8: An impact pathway, in this case for climate services



If productivity has increased or costs reduced, this could result in increased income. This income could be reinvested into other resilience-building activities or priorities, such as education and healthcare. Based on positive results along the impact pathway, a community institution may increase its leverage to advocate for better public services, and so on. The better these different levels of impact can be captured and used to support learning and management, the more effective the intervention stakeholders will be in managing for impact. Key challenges moving along the impact pathway include:

- the further down the pathway an assessment progresses, the harder attribution becomes. This is largely due to other variables beyond the control of the project exerting progressively more influence on the impact being measured, diluting the effects that can be attributed exclusively or even partly to a project activity
- impact becomes harder to anticipate and therefore develop indicators for. This puts an emphasis on establishing systems for ongoing, continuous data collection (especially by participating communities) and the flexibility to incorporate indicators that participating communities and households define, rather than occasional, large-scale extractive exercises measuring externally defined indicators of limited relevance.

2.5 Time and space considerations – the impact catchment

The consequences of resilience-building measures across time and space are important parts of understanding impact. Assessment should at least seek to consider:

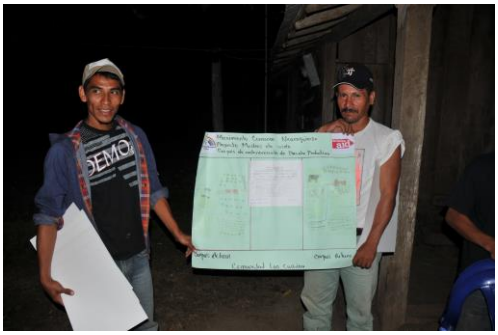
a. whether an intervention increasing resilience in the short-term could lead to greater vulnerability in the longer term

b. whether an intervention increasing the resilience of one area or community could result in increased vulnerability in another.

The former emphasises the strong interrelationships between **increased resilience and increased sustainability** that does not discount the future, but seeks to enhance assets and their buffers – an approach particularly well-developed with respect to environmental sustainability, thus demonstrating the importance of this interconnection with resilience.

The latter highlights the need to view resilience from a catchment perspective and ensure that impact assessment includes direct, indirect and unintended participants across the **impact catchment**. For example, a flood protection embankment that protects one village from severe flooding by moving floodwaters downstream into the next village or town benefits the inhabitants of the first, but only by displacing the risk onto other communities. Likewise the vulnerability of both may be gradually increased by long-term degradation and deforestation further up the catchment. The impact catchment approach shows how resilience impact assessment needs to be more three dimensional, avoiding a reduction to simplistic, narrow 'number of direct beneficiaries' or 'value for money'-type measures of impact.[14]

Below: Community members in Nicaragua explain their resilience action plans, highlighting the impact pathways they anticipate as a result of the activities they have prioritised.



Below: Risks affect different categories in society in different ways – this highlights the importance of involving all parts of a community in a resilience impact assessment. Here women across three generations of a community in Tajikistan discuss climate change, resilience and marketing and the priorities they see as key to addressing these issues.

2.6 Gender, inclusion and ethics

A number of characteristics common to all populations have a fundamental impact on vulnerability to risk and therefore should be addressed in all impact assessment activities. As risks tend to affect women and men in different ways, especially where the division of labour in a household or community is more pronounced, gender is of particular importance. In an assessment, this means ensuring that women and men have space to answer questions about impact as freely, accurately and objectively as possible, for example, through separating focus group discussions into groups of men and groups of women. For this to be operationalised effectively, an impact assessment team should also be gender-balanced, so that women responders can be facilitated by a woman impact assessor and men by a male impact assessor.

Where *a priori* reason exists that suggests, for example, a certain risk is more relevant to women (for example, as per their PVCA analysis), this should be reflected in the methodology. Likewise, impact evidence should be triangulated between women and men to ensure that gender biases do not push the findings in one direction or the other, undermining their validity. Ordinarily, the sample size should demonstrate equality between men and women to ensure that the concerns and priorities of both are reflected equally in the final impact assessment. Gender disaggregation is valid even when responses are on behalf of the household, as it can highlight and differentiate the resilience priorities of women and men.

Risk also affects different age groups in different ways, indicating that stratification of a sample by age group may also be important. Elderly responders have particular value due to their long experience of the shocks and stresses that can affect a programme area. This provides valuable, contextual information that can be used to triangulate more formally-collected or quantitative data. Other categories that are relevant to measuring resilience impact include the disabled and livelihood groups, such as agriculture and fisheries, engaged in sectors that are likely to be relatively more vulnerable.

Ethical practices should always be followed, with particular sensitivity in resilience measurement to the more controversial shocks such as conflict or domestic violence. Basic principles include informed consent, anonymising responses, declaring any conflicts of interest, considering the safety of enumerators and respondents at all times and ensuring that the results of assessment are fed back to all stakeholders. Further guidance can be found in the Social Research Association's *Ethical Guidelines* (see Annex 1, section B).

3. Characteristics and indicators for resilience impact assessment

When measuring the impact of an intervention that seeks to increase resilience, there are a number of features that distinguish indicators focusing on this from those used in a conventional livelihoods approach. This is not to say that existing impact assessment methodologies are irrelevant. On the contrary, measuring resilience should build on the wide body of practice that has already proved effective.

3.1 Indicator characteristics

So resilience indicators should follow the widely accepted basic principles of being SMART:

- **S**pecific – not waffley or vague
- **M**easurable – can be either quantified or described qualitatively
- **A**chievable – realistic given the resources available
- **R**elevant – in this case, actually describe improved resilience to risk rather than just general livelihood improvements
- **T**imebound – achievable within a defined period.

In addition, they need to describe a change in the livelihoods of those involved in the intervention, rather than just a result, i.e. impact rather than output. So running 20 resilience seminars for communities is an output – it only has relevance to impact if this training leads to a specific improvements in resilience that can be (a) **detected** (or measured) to show significant change, (b) **triangulated** (or confirmed from more than one source of information) to confirm the reliability of this detection, and (c) **attributed to** (or shown convincingly to be at least partly resulting from the activity, in this case the 20 resilience seminars).

Other characteristics of resilience indicators that are different from business-as-usual include:

- a focus on anticipating and managing risk related to either specific events (a shock) or incremental stress
- a focus on buffer resources and how these have changed in managing a risk or stress (as per section 2 above)
- an understanding that livelihoods are dynamic in moving forward and backwards in response to shock and stress, and that positive resilience impact does not necessarily mean short-term progress at all times
- a focus on how assets recover effectively in response to mitigation of a shock or stress
- a focus on how livelihoods change and/or diversify in response to shocks and incremental stresses, increasing robustness
- due to the context specificity of resilience, failure to develop (at least some) resilience indicators directly from the intervention participant/community perspective will result in under-assessment (see 3.2 below) or even miss entire categories of impact

‘Resilience is a process, not a trait. It is not enough to identify protective factors, because these do not create resilience in all cases. Resilience is created when these factors initiate certain processes’

Professor Michael Rutter, King’s College London

- the importance of damage avoided and understanding the difficulty of determining positive impact based on measuring something that has not happened, rather than something that has
- linked to this is the importance of resilience as a way of reducing input costs and increasing the efficiency of processes, often as or more important than simply increasing output productivity
- the need to regularly review and revise indicators as the risks of shocks and stresses, and the information that is available on these, change.

Some qualitative indicators that can appear difficult to measure numerically at first may become quantifiable once the impact is understood more clearly. So with a PVCA-based approach that, for example, seeks to increase access to social protection public services in India (such as the employment guarantee and pre-school child centres), increased empowerment, planning capacity and collective action can be monitored financially through logging the value of services that have been successfully obtained by each community over time and then aggregating this data to get a total value, giving some indication of the financial returns to the activity.

Below: Community members from a village in Uttar Pradesh, India, explain the map they have developed as part of a resilience-focused approach, highlighting households registered to receive weather forecasts.

3.2 The importance of context specific indicators

When indicators are used as specific examples of change, often with the emphasis on qualitative impact, different characteristics become important. Section 3.3 (below) shows 20 generic indicators (albeit based on actual examples): while these are useful in giving some general guidance on what a resilience indicator looks like, the context specificity of resilience impact requires a recognition of indicators that are generated by the communities or participants involved. Externally defined indicators will often not be the same as those defined by the project participants or communities, may not focus on measuring impact that the participants really value, and are more likely to miss unanticipated impact. In this context, we can refer to **SPICED** indicators:[15]

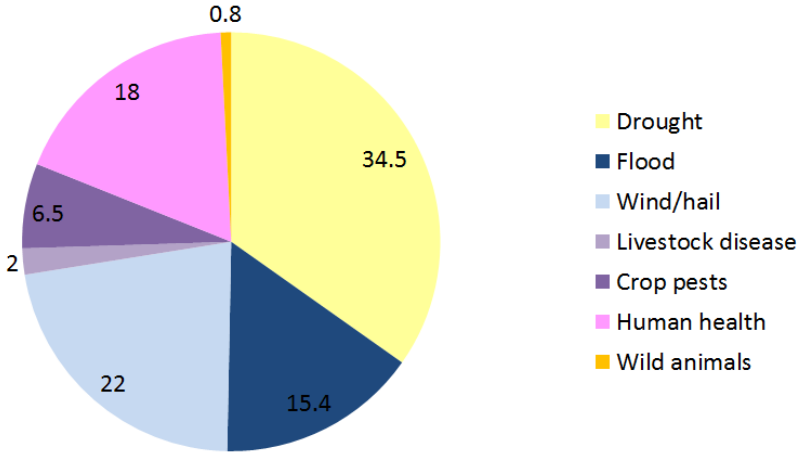
- **S**ubjective – informants have a special position or experience that gives them unique insights that may yield a very high return on the investigators' time. In this sense, what may be seen by others as 'anecdotal' becomes critical data because of the source's value as a key informant
- **P**articipatory – indicators should be developed together with those best placed to assess them. This means involving a project's ultimate participants and beneficiaries, but it can also mean involving local staff and other stakeholders
- **I**nterpreted and communicable – locally defined indicators may not mean much to other stakeholders, so they often need to be explained
- **C**ross-checked and compared – the validity of assessment needs to be cross-checked by comparing different indicators and progress, and by using different informants, methods and researchers
- **E**mpowering – the process of setting and assessing indicators should be empowering in itself and allow groups and individuals

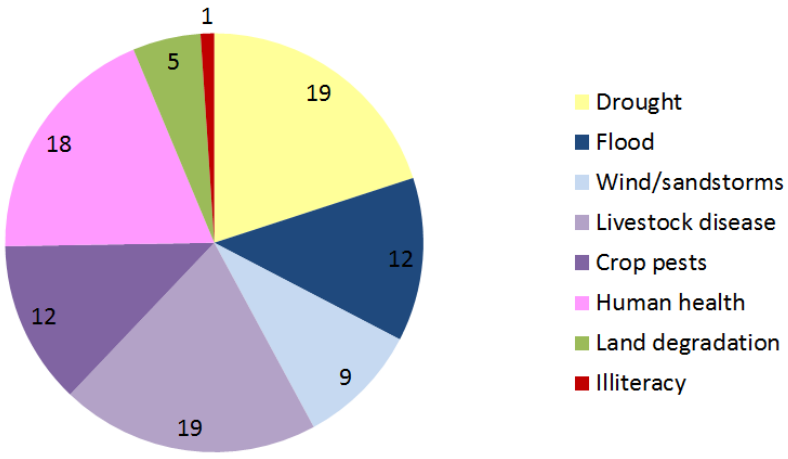
to reflect critically on their changing situation, building the capacity of participants to measure their own impact effectively

- **D**iverse and disaggregated – there should be a deliberate effort to seek out different indicators from a range of groups, especially men and women. This information needs to be recorded in such a way that these differences can be assessed over time.

The most important factor is ensuring that the indicators developed to measure resilience are therefore context relevant and based on the risk profile that the community or area has agreed.

Figure 9: Community risk profiles from Malawi (top) and Burkina Faso (bottom)[16]





This will ensure that, as far as possible, indicators are both SMART and SPICED – focused on the risk-reducing and resilience-building factors that are relevant to the local context and prioritised by vulnerable communities. This is particularly important in the use of proxy indicators, where measurement is indirect and usually based on an external 'expert' view that the correlation between the impact being measured and proxy indicator is reliable. Both more generally and particularly with resilience, this correlation can be tenuous. Use of proxy indicators is therefore only valid as a last resort and where the justification is especially strong.[17] Most impact is best measured directly.

3.3 Typical resilience indicators

This is not an exhaustive list, nor does it imply that a resilience project should seek to measure these indicators. It is a menu that may be useful in providing guidance to intervention planners and evaluators on the kind of indicators they should be looking at, given the characteristics of resilience, as described above. Actual indicators should always be based on the resilience context being measured and the resilience priorities, as expressed by those participating in a resilience-building intervention.[18] See Annex 2 for assessment tools (also divided into quantitative and qualitative) to match to these indicators.

'There is nothing like first-hand evidence'

Sherlock Holmes

Figure 10a: Typical **quantitative** resilience indicators

Indicator	Key measurables	Example[19]	Main elements of resilience
1. The increase in warning time that an operational early warning system gives to those receiving households and communities.	○ **Warning time in hours/days/weeks** after establishing an early warning systems (EWS) vs warning time before.	○ In the Sundarbans, communities on the Indian side received only two hours warning of Cyclone Ayla whereas those on the Bangladesh side received 24 hours.	○ Adaptive capacity
2. The lives and livelihood assets saved as a result of increased warning times and access to safe shelter.	○ **No of lives/value of assets saved** after EWS establishment or improvement vs before. ○ **No of people accessing safe refuge** after establishment of shelters vs before as a proportion of the population.	○ As a result, while 600,000 people were safely evacuated in Bangladesh, 400,000 were stranded in India.	○ Buffers
3. The speed at to which a livelihood can recover to full capacity after a significant short-term shock, e.g. cyclone, earthquake, conflict.	○ **Recovery time in days/weeks/months** of communities involved in resilient livelihoods programming vs before; trend in recovery speeds (in the case of repeated shocks). ○ **Reduction in asset loss** – lower asset degradation by a disaster event or incremental stress, e.g., of sustainable agriculture vs conventional agriculture, such as through reduced land degradation, crop destruction, etc.	○ Farmers in Central America recovered faster post-Hurricane Mitch if they practiced organic techniques. They lost less soil and, due to a greater use of erosion control and organic techniques, were able to resume production faster than their non-organic neighbours. Specifically after Mitch, they had 40% more topsoil, lost 18% less arable land to landslides and had a 49% lower incidence of landslides than conventional plots.	○ Recovery ○ Robustness
4. The degree to which a livelihood has diversified and thereby spread risk across a range of livelihood options, including measures that either protect existing livelihoods from increased climate and other risks or transform them into new lower-risk enterprises.	○ **Number of livelihood operations** increasing resilience successfully implemented after support vs before (e.g. protection – increase in number of drought-resilient crops or crop varieties grown; and transformation – increase in number of new, less vulnerable enterprises established). ○ **Increase in medium – long-term overall income** after diversification vs before.	○ Farmer in Uttar Pradesh, India, practising sustainable agriculture indicated that they grew up to 27 crop varieties as opposed to two or three when they were chemical farmers. They assessed this as both more resilient to changing weather and more profitable in local markets than their previous practices.	○ Buffers ○ Thriving

5. The increase in diversity of marketing options and/or markets that reduce reliance on one or only a few vulnerable production options and/or relatively low-value markets.	○ **Number of markets/marketing options accessed** after vs before. ○ **Increase in prices achieved per unit of output,** returns to marketing.	○ Farmers in northern Tanzania increased their marketing options through growing seed maize, vegetable crops and establishing credit schemes, which together with increased collective marketing, increased the prices gained by 10% and reduced reliance on low-profit maize sales to a small group of local traders.	○ Thriving ○ Buffers
6. Risk information, such as seasonal climate forecasts, short-term weather forecasts, and market forecasts, has been used to make operational livelihood decisions that result in increased output and income.	○ **Sustained increase in units of output** (e.g., yield per hectare); increase in income per unit of input (e.g., gross margin per hectare) of those using risk management information vs the average for the location and/or yields achieved before.	○ Farmers in Zimbabwe that gained increased access to seasonal forecasts and used them to make pre-season decisions on crops to grow, when to plant, etc., experienced a 9-18% increase in yield vs those with little or no access.	○ Adaptive capacity ○ Thriving
7. The ability of a livelihood to manage and absorb small – medium stresses caused by short – medium-term variation.	○ **Differential in units of output** of those receiving support vs the average for the location and/or yields achieved before. ○ **Reduced cost of inputs required,** including labour and ecosystem services, e.g. clean water. ○ **Damage avoided** as a result of forecast information, protective infrastructure, etc.	○ Farmer groups in Kenya highlighted the importance of seven-day forecasts in managing dry spells or intensive rainfall episodes through improving the timing of pest control, fertiliser application and harvest timing, contributing to an increase in yield of more than 15% for two-thirds of those receiving the forecast via SMS. ○ Farmers in India using forecasts highlighted reduced input cost as the most important impact, citing up to 50% reductions through better timing of irrigation, planting and fertiliser (compost) use.	○ Robustness ○ Buffers ○ Adaptive capacity
8. The ability of a livelihood to manage and absorb ongoing stresses caused by incremental change.	○ **Differential in productivity** of those receiving support vs before. ○ **Change in returns** of new enterprise(s) vs those previously practised.	○ As a result of sea level rise and increased soil salination, communities in the Sundarbans of Bangladesh have switched from rice to shrimp farming with a consequent increase of 30% in income per hectare.	○ Buffers ○ Robustness
9. Buffers that mitigate risk of both shocks and stress are enhanced.	○ **Buffers are moving towards an ideal level (the optimal level of redundancy)** in their ability to mitigate risk and therefore reduce the costs of future shocks and stresses to financial, physical and natural assets.	○ In Bangladesh, a network of 2,500 cyclone shelters buffer the population against cyclone risks, reducing loss of assets and facilitating a quick return to recovery and long-term resilience building. The cost-benefit ratio of measures such as these returns US$ 3-27 to every $1 of cost.	○ Buffers ○ Recovery
10. Use of forecasts enables prediction of disease outbreak thereby reducing the spread of disease through pre-emptive actions.	○ **Reduced numbers of infected** persons as a result of pre-emptive measures. ○ **Reduced cost of treatment** due to pre-emptive measures as compared to previous comparable years or to communities not taking measures.	○ In Kenya, the Kenya Medical Research Institute has worked with the Kenya Met Dept. to investigate the possibility of using forecast models to pre-empt malaria outbreaks. This has the potential to reduce both infection and cost of treatment.	○ Adaptive capacity

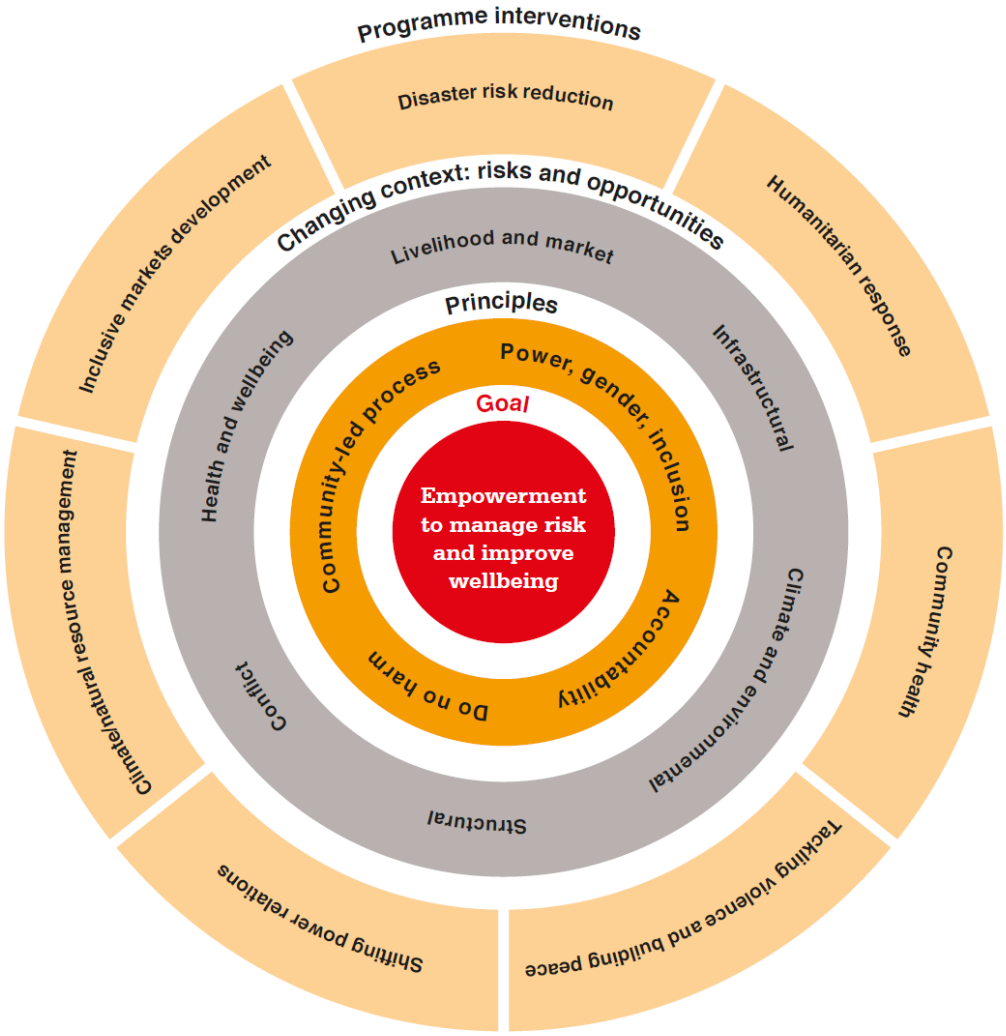
Figure 10b: Typical **qualitative** resilience indicators

Indicator	Key measurables	Example[20]	Main elements of resilience
11. The capacity of individuals and their communities to both access and understand technical information on risks, e.g. climate science.	○ **Increased analytical capacity** at household and/or community level.	○ Farmers in central Kenya increased their understanding of, and their ability to navigate, the uncertainty in scientific forecasts.	○ Adaptive capacity
12. The capacity of individuals and their communities to combine local knowledge with technical information on risks, e.g. geological surveys with local experience of earthquakes; climate science with local bio-indicators used to anticipate weather.	○ **Increased capacity to combine local information with technical information** at either or at both household and community level; increased understanding of external risk information and its relevance to the local context; increased responsiveness of external technical information providers to local resilience priorities.	○ These farmer groups combined the scientific forecast with about 10 key local indicators, the latter mainly focused on the critical decision of when to plant.	○ Adaptive capacity
13. The capacity of individuals and their communities to use this analysis to make key resilience-building livelihood decisions.	○ **Use of information to make specific decisions** – specifically either new livelihood decisions that have been made or existing ones that have changed as a result of increased analytical capacity and access to information at household and/or community level.	○ Forecast users highlighted several key decisions that they had made differently as a result of understanding forecasts better, including when to plant, what crops/varieties to choose and how to manage land for greater moisture conservation; as well as domestic resilience-building decisions, e.g. food storage decisions.	○ Adaptive capacity ○ Robustness ○ Buffers ○ Thriving
14. The capacity of communities to develop adaptation plans that correctly identify key risks and those in the community most vulnerable to them, and propose concrete measures that can be put into action.	○ **Increased planning capacity**; increased social capital and collective risk management; increases in the quality of community-based planning using short-, medium- and long-term climate and other risk information; increased access to support (technical advice, investment, or both).	○ In northern India, 124 communities developed community action plans as a result of participatory vulnerability and capacity assessment (PVCA) processes, highlighting increased levels of collective action, especially in advocating to local government for services they were entitled to but not receiving (such as employment on public works, child development services, etc.)	○ Adaptive capacity ○ Buffers ○ Thriving
15. The capacity of individuals and communities to manage livelihood measures that result in an increase in resilience to short-, medium- and long-term change.	○ **Increased management capacity**; increased social capital; plans applied to specific resilience-building measures; **increased accountability of local government** to communities.	○ In Burkina Faso, adaptation training enabled communities to understand climate risk more clearly, analyse their situation over the past 50 years and start developing community maps that highlighted areas of flood risks, areas to enhance reforestation, etc.	○ Robustness ○ Buffers ○ Adaptive capacity
16. Improved planning at multi-community/ catchment level reduces vulnerability to short-, medium- and long-term risks and increases the accountability and responsiveness of governance structures.	○ **Increased technical capacity** to assess upstream and downstream risk and vulnerability; increased ability of different stakeholders to coordinate planning; **increased accountability of local government** to communities, especially in **integrating community-based plans into local**	○ In the Eastern Cape of South Africa, community groups that had developed participatory land use plans succeeded in getting their plans formally integrated into local District plans and also in drawing down specific financial support to implement key aspects of these plans.	○ Thriving ○ Buffers ○ Robustness

	government plans.		
17. Buffers that mitigate risk of both shocks and stress are enhanced.	○ **Buffers are moving towards an ideal level (the optimal level of redundancy)** in their ability to mitigate risk and therefore reduce the costs of future shocks and stresses to human, social and natural assets.	○ Shocks: early warning of cyclone Phailin reduced mortality to 45 deaths compared to more than 9,800 during Cyclone 05B in 1999 in almost exactly the same place (coastal Orissa in eastern India). ○ Stresses: Farmers in Kenya strengthened their group and individual capacity to make climate-smart decisions through access to forecasts, reducing their exposure to planting failure and drought stress, protecting soils through conservation agriculture, and increasing yields (96% reported +5% or more impact on yield).[21]	○ Buffers ○ Adaptive capacity
18. Health systems are able to adapt and respond to either emerging health threats (e.g. Ebola) or to the shifting needs of the population through better integration.	○ **Enhanced health surveillance, early warning and capacity to respond to health concerns and disease outbreaks** facilitates early response to minimise negative impact, reducing impact on lives and livelihoods; increased effectiveness of investment in health services and health resilience; reduced response cost.	○ Dynamic coordinated responses are improved through integration of previously separate and parallel services for different health issues (e.g. HIV, TB, maternal and child health, etc.). Early warning systems used for other shocks and stresses include health risks.	○ Adaptive capacity ○ Buffers
19. Greater understanding of individuals and communities gives them the ability and motivation to access, understand and use information supporting good health, leading to positive health seeking and promoting behaviour.	○ **Increased public health awareness reduces the risk and impact of illness**; reduced risk and level of catastrophic expenditure by those affected.	○ Public health awareness of, for example, water-borne diseases such as cholera and ways to avoid infection result in avoided illness and reduces the risk of households cutting down on other necessities, such as household nutrition or school fees, or the affected individual not accessing the services they need, with potential negative impacts on individuals, households and livelihoods.	○ Adaptive capacity
20. Increased social cohesion reduces the risk of intra-community tensions, facilitates discussion of contentious issues and dispute resolution processes.	○ **Increased social capital in addressing tension and conflict within the community;** greater openness and confidence in resolving vulnerability issues that were previously difficult or impossible to expose and discuss.	○ Women's groups in West Bengal, India were empowered by participatory vulnerability and capacity assessment (PVCA) processes to address domestic violence and its underlying causes, taking specific actions around improving gender relations, the increased use of migration to access employment, and the proliferation of illegal alcohol retailing.	○ Buffers

3.4 How resilience indicators fit with the Resilience Framework

Figure 11: Christian Aid's Resilience Framework [22]



The Resilience Framework highlights seven key programme intervention areas. The relevant resilience indicators related to these are therefore (as per the indicator number in section 3.3 above, there is overlap, e.g. early warning can relate as much to conflict as to cyclones, but the main connections are highlighted here).

- **Shifting power relations**: achieving just power relations from household to national levels by building an environment that enables participatory decision making, representation and accountability through inclusive and consultative processes: **14; 16.**
- **Climate change and natural resource management**: sustainable management of land, water, soils, forests and agriculture that reduces poverty and hunger in the face of climate change, improving the resources a community depends on for future generations: **7; 8; 11; 12; 13.**
- **Inclusive markets development**: supporting producers to gain increased control over and access to information, credit and markets so that they can build sustainable livelihoods, increase income and create savings: **4; 5; 6.**

'At the core of our framework is our belief that individual and community resilience can be enhanced by empowering poor and vulnerable women and men, boys and girls to manage risks and improve their wellbeing, so that they can live with dignity'

- **Disaster risk reduction**: reducing disaster risks through systematic efforts to analyse and manage the causal factors of disasters, including through reduced exposure to hazards, lessened vulnerability of people and property, wise management of land and the environment, and improved preparedness for adverse events: **1; 2; 10; 15; 17.**
- **Humanitarian response**: partners and communities are organised and empowered to implement the functional aspects of managing emergencies whenever they arise: **3.**
- **Community health**: strengthening access to services that enable people to maintain healthy and productive lives and provide nutrition and environmental health management that promotes health and wellbeing: **9; 18; 19.**
- **Tackling violence, building peace**: increased protection for those most vulnerable to violence, and equipping them to address the causes of violence, to tackle impunity, to develop peaceful and effective alternatives to violence and armed conflict: **20.**

Below: Making technical frameworks relevant and useful to local resilience planning, implementation and assessment is key to bridging the gap between the sometimes more esoteric concerns of development experts and the realities of risk reduction. Here communities in Nicaragua together with their local civil society organisations have adapted the asset pentagram from the sustainable livelihoods approach into a locally appropriate format to integrate its concepts into their resilience planning.

4. Key issues in implementing impact assessment

4.1 Detection, triangulation, attribution and contribution

Detection, attribution and triangulation are key concepts in ensuring that resilience impact has been reliably identified and demonstrated to be as a result of an intervention and not some other factor, such as an alternative intervention, a government programme, or just a general improvement in economic or environmental conditions.

- **Detection**: the determination of an impact that may or may not be due to an intervention.
- **Attribution and contribution**: attribution is the establishment of a cause and effect relationship between an intervention and the impact measured. Contribution is essentially the same, with the distinction that where attribution generally implies the effect is largely due to the cause, contribution is used where the link between an intervention and the impact measured is not due to the intervention alone but other factors as well. It can also be referred to as partial attribution.
- **Triangulation**: involves the comparison of data using more than one assessment method or more than one source of information, to assess the impact of the same activity and output. It reduces bias and increases validity and reliability.

There is increased appreciation of the need to understand the various factors that contribute to impact, and whether or not they are the result of any particular intervention. For example, crop yields will change as a result of several factors, just one of which may be the implementation of a new drought-resistant crop variety through a resilience-building project. The challenge for impact assessment here will be to demonstrate the significant contribution of the new crop variety, rather than attempting to measure the precise attribution of impact to the intervention.[23]

These three processes are important in strengthening the **validity** of the evidence – reducing the effect of biases (see section 4.6 below), ensuring as far as possible that the impact evidence is genuine for the population or area being measured (internal validity), that it doesn't change when measured by more than one approach (convergent validity), and that it is applicable to other, similar interventions (external validity).

Reliability of the evidence is strengthened by using the most appropriate measuring tools (see Annex 2), sample size and sampling techniques (see 4.3 and 4.4 below).

In particular, where **triangulation** between qualitative and quantitative data (and vice versa) can be used to confirm the accuracy of impact data, this increases confidence in impact results – given the additional challenges of measuring resilience impact mentioned above (in 2.3 and 2.4), the application of mixed quantitative/qualitative assessment methods is even more important. Without this sort of triangulation, statistical measures of impact are particularly vulnerable to inaccuracy and manipulation, resulting in significant over- or under-assessment of impact.

'Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted'

Albert Einstein

Likewise, purely qualitative measures to infer overall impact of an intervention that are not cross-referenced with quantitative measurement are vulnerable to biases, which undermine the validity of the information. So, for example, using a case study of a single or small number of project participants to represent overall impact is unreliable, without more representative statistical evidence confirming that the case study does in fact show the typical impact across a larger number of participants, and invalid without triangulation with another source or measurement approach.

Below: Communities in Kenya explain to each-other the various risks they face and their scoring on a percentage scale before and after the implementation of resilience-building activities. This and other forms of self-assessment are a valuable part of measuring resilience impact.

4.2 Resilience baselines

The basic objective of a baseline is to establish a set of benchmarks against which future progress can be measured, usually at the beginning of an intervention.

The limitations of using control groups in resilience programming (see 4.5 below) puts emphasis on establishing a good baseline, but this should not leave their problems unacknowledged – baselines frequently lack relevant data when impact assessment is being undertaken several years later due to the adjustment of the project to circumstances during its implementation and the inevitability of impacts that were unanticipated at the beginning. The lack of a baseline does not necessarily mean that impact cannot be measured. Many assessment tools (see Annex 2) can be used retrospectively to generate valid impact evidence.

Other challenges also arise. The standard baseline approach is to measure as much as possible about the project environment and population at the beginning of an intervention, often focusing on dubious proxy measures such as the 'number of meals eaten per day', in the hope that there will be enough information about the pre-intervention situation recorded to demonstrate some positive impact by the end of the intervention. This catch-all approach typically ends up:

- confusing a general situation analysis with a baseline
- measuring the wrong things
- only measuring useful benchmarks by accident, but then burying them in irrelevant detail
- generating large amounts of data that causes 'analysis paralysis', resulting in a poor quality report or delaying its finalisation
- absorbing a disproportionate amount of the impact assessment budget, leaving little for important further continuous or adhoc assessment
- requiring expensive outside consultancy support to implement.

Even well-executed baselines typically try to measure too much and therefore fail to focus on benchmarking priority indicators in which project implementers and participants have **a strong expectation of subsequently seeing impact that can be attributed to the intervention.**

Baseline questions should also pass the **repeatability test** – i.e. if the question is asked again at the end of the intervention, is it likely to generate a response that will provide good evidence of impact?

A strong resilience baseline should:

- focus on measuring a small number of priority indicators that the project expects to see significant change in by the end of implementation, as a result of that implementation
- then maintain strict discipline, to avoid 'mission creep' by refusing to allow the addition of questions on things that would be 'nice-to-know' but not 'need-to-know' or that someone in head office has dictated should be measured because 'it's an organisational strategy priority'
- make sure that these indicators are agreed by implementers and participating households and communities
- make sure that these indicators focus on the risk profile of the population being measured
- make sure that the data can be easily processed and documented in accessible ways (graphs, pie charts, etc.)
- put in place the measures that will facilitate the smooth start-up of key continuous data collection that will deliver rich and strategic information on impact, including unanticipated impact, as it emerges from project activities and outputs
- above all, keep it simple and avoid creating unnecessary data collection and processing burdens.

The limitations of baselines in terms of unanticipated impact are obvious. Shocks tend to occur intermittently, and possibly not at all, during the life of the project. With community-based processes using action plans, impacts will vary from community to community depending on the risks prioritised in their action plans, so a baseline survey may lack the flexibility to establish benchmarks that are valid across the project area. Resilience impact assessment therefore relies as much on continuous assessment and retrospective analysis as it does on baselines.

Below: Community group leaders in West Bengal, India review their rainfall data. With the growth of citizen science and other bottom-up approaches to data gathering and analysis, communities' own data becomes a valuable source of evidence for planning, implementation and impact assessment.

4.3 Sample sizes

Impact assessment processes, especially at the baseline stage, can become preoccupied with sample sizes. However, it is **representativeness** that is the more important issue rather than any particular number. Given the difficulties of using control groups for assessing resilience impacts (see 4.5 below), this puts more emphasis on baseline data and therefore the importance of ensuring this is as useful as possible in setting the initial benchmark.

Sample groups should ensure that baseline exercises are manageable and collect high quality (as opposed to high quantity) information. The main concerns are identifying a trend in impact (and therefore an intervention's contribution to impact) rather than coming to a definitively precise measure. Sample group sizes can be limited to reflect this and keep the exercise manageable in respect of the:

- precision of the estimate
- accuracy and reliability of measurement
- degree of cross-tabulation anticipated
- ensuring a representative sample through selection
- cost of the process

rather than arbitrary concerns about statistical significance. This implies smaller rather than larger sample group sizes.

Figure 12: Setting the right sample size

The first table of numbers shows the relationship between sample size and **margin of error** for a large population where the cut-off for a +/- 5% level occurs below 500.

The second table shows how this relates to population size where assessment requires a 95% **confidence level**. The general conclusion is that for impact assessment purposes (which does not need academic research levels of statistical confidence), sample sizes above 400 add no extra value. Integrating effective triangulation suggests that this figure can be reduced further without any loss of confidence in the impact measured (although views differ on exactly how much reduction is possible). In qualitative assessment where samples are typically much smaller, the concept of **saturation** is useful. This suggests that where increasing the sample size with additional respondents does not improve explanations or add significant detail to the assessment, this is where the sample should stop.

Sample size	10	20	50	100	200	500	1,000	2,000	5,000	10,000
Margin of error	31.5	22.4	14.1	10.0	7.1	4.5	3.2	2.2	1.4	1.0

Sample sizes needed for a 95% confidence level with a margin of error of +/- 5% [24]

Population size	**Sample size without triangulation**
10	10
100	80
200	132
300	169
400	197
500	218
600	235
700	249
800	260
900	270
1,000	278
2,000	323
5,000	357
10,000	370
20,000	377
100,000	383
1,000,000	384

Establishing a working equation means:

a. **P**recision – margins of error of 5% are adequate for impact assessment, a level reached at sample sizes between 200 and 500. To achieve margins of error of only 1% (more typical of in-depth formal research or research science levels of significance), a sample of 10,000 is needed. This small increase therefore implies a cost increase of 20x the level needed to obtain easily adequate results, not a constructive use of scarce impact assessment resources. So in general, for impact assessment, samples of more than 400 rarely add greater value. Likewise a confidence level of 95% is almost always quite acceptable.

b. **A**ccuracy – or quality of data. The larger the sample size, the greater the likelihood that the process turns into a substantial and challenging logistical exercise. As these challenges are encountered, interviewing each household or individual becomes an increasingly difficult task. The temptation to cut corners and treat the gathering of information as an increasingly robotic or tick-box process erodes the quality of the data.

On the other hand, keeping samples manageable and using mixed methods to systematically triangulate the information collected suggests smaller samples.

c. **R**epresentative samples – usually means a form of random sampling. This may include stratifying the sample, for example, according to agro-ecological zone or clustering the sample according to social characteristics. Different strata or clusters will need an adequate sample for each, which may push the sample size up. Here again, there have to be very clear reasons as to how stratification or clustering adds to the value of the data and why this added level of detail is needed. One good reason is that it is directly related to the level of cross tabulation needed.

d. **C**ost – impact assessment resources are typically categorised as overheads and therefore subject to intense scrutiny to justify their expenditure. Strict prioritisation is therefore needed in terms of effectively measuring a smaller number of key variables that the project expects to affect strongly (as per the indicators developed), rather than measuring a large number of variables less well that may be strongly or weakly affected by the project, e.g. using general livelihood surveys for baselines. One benefit of ensuring community involvement in setting indicators (and therefore what to include in the baseline) is that this can enable the development of an information collection process that groups and group leaders can participate in managing, thereby increasing the cost-effectiveness of assessment and empowering participants' own assessment capacity.

e. **S**ub-divisions for cross tabulation – if several sub-divisions are being presented, e.g. the results of crop yields for 4 different holding size categories and for men and women farmers, a representative sample in each division is needed. This may push the sample size up, but usually only where large numbers of variables are being cross-tabulated. If this is the case, questions arise as to why so many variables are needed so, for example, in a process where 16 holding size categories are described, is it really essential to break the sample down to 16 sub-divisions when two or three might be perfectly adequate?

The confidence in the impact revealed lies in high-quality information-gathering processes and strong attribution and triangulation, rather than bidding up survey sample sizes under the mistaken assumption that this adds value to the assessment or baseline process.

A common question is: *'Is there an equation that we can use to determine sample size?'* The short answer is no, not in the numerical sense. But as sample size is a function of (***ƒ***) the key elements to consider (as above), a functional equation would be:

Sample size = *ƒ* (P, A, R, C, S)

Below: Community leaders in El Salvador explain early warning and flood management systems to evaluators. While small-scale farmers have been active in addressing flood risks, the behaviour of commercial sugar estates further up the catchment has created challenges to their resilience. This illustrates the importance of understanding the impact catchment in order to measure how resilience is or is not enhanced.

Figure 13: Setting the right sample size

	P	A	R	C	S
Sample size = ***f***	**Precision** or margin of error needed, usually 90-95%. This is the starting point and the numbers of people included will be between 100 and 400. Note: participatory methods use smaller samples; formal surveys tend to go for larger ones.	**Accuracy** or quality of data. Larger samples usually mean lower-quality data – enumerators get tired towards the end of the day, data collection efficiency varies between enumerators, etc. This is especially true when with open-ended questions are used.	The method needed to ensure the sample is **representative** of the population being measured, e.g. through random sampling. It is more important that the sample is representative than it is of any particular size.	**Cost**, or what budget the project can afford, and **capacity**, or what the implementers can manage within the budget most successfully. This should take into account the number of methods to be used for impact assessment and the need to process the data into a final report that delivers quality information on time. Triangulating survey data (i.e. measuring the same thing with more than one method) will improve its validity, and effective triangulation suggests that a more manageable sample size can be used.	The number of **sub-divisions** or stratifications in the sample, e.g. if you want to look at responses over four different farm holding sizes, your sample needs to be the right size for a representative sample in each of the four categories.
Temptations to avoid, issues to focus on:	Despite widespread perceptions and temptations to bid up sample sizes, there is little gain in precision above this range. External consultants seem to have a tendency to bid up sample sizes (maybe because it means more work for them).	Data quality and consistency tend to be undermined by large samples, not enhanced – and it is expensive to repeat interview schedules when poor quality data has been gathered.	It is not the size of the sample but the way the population is sampled that should be the focus – small samples can be defended if the arguments for their being representative are strong.	Impact assessment resources are 'overheads' so tend to be less than required – this means making careful decisions about how to get convincing evidence that uses quantitative data to triangulate qualitative and visa-versa. The challenges of processing the data into a quality report are almost always underestimated – processing large sample surveys can be especially complex.	Sample size is <u>not</u> **P** x **S**

4.4 Sampling techniques

As Figure 13 implies, ensuring that the sample is representative of the population being measured is the key priority. Sampling methods are a key part of this process, and fall into either random or non-random categories. Random sampling techniques include:[25]

- **simple random sampling**: a group of individuals or households are selected from a numbered list of all individuals or households at random using random number tables or simply putting all the individual or household names on bits of paper in a bag and pulling out the required number for the sample. This is a useful technique if the project area or the population is relatively uniform in character
- **stratified random sampling**: the overall population is broken down into sub-groups, which are then sampled. The sample should reflect the relative size of each group, for example, if the stratification is based on three agro-ecological zones and the population is divided 50%, 30% and 20% between these zones 1-3, 50% of the sample should be selected from zone 1, 30% from zone 2 and 20% from zone 3. This is a useful approach to projects that have varied characteristics
- **cluster sampling**: the project area is divided into clusters, for example, groups of villages and sampling occurs in two stages. First a random sample of clusters is made and then a random sample of individuals in each cluster. This can reduce the cost of data collection as it is limited to a less scattered sample
- **staged sampling**: similar to cluster sampling but adds a third stage, for example, sampling a number of sub-districts before sampling villages and then individuals
- **random walk**: where information on the sampling frame is incomplete, it may be easiest to use a random walk through the community, for example, interviewing every fifth house if you want to cover 20% of the community in a particular sample. Like simple random sampling, this works best with relatively uniform areas or populations.

Non-random sampling includes:

- **quota sampling**: a number of respondents are selected based on categories, such as gender, based on first come first served, or the easiest to reach given the time available. This can open the sample up to biases, such as tarmac bias (see 4.6 below)
- **chain sampling**: often used in qualitative interviews where each respondent then suggests the next appropriate interviewee.

In practice, the first three methods will cover the majority of situations found in most project areas.

Below: A small group of villagers assess their flood protection embankment in Bangladesh. The resilience benefits that this sort of infrastructure provides will likely affect different people in different ways (depending on livelihood, location, etc.), highlighting the importance of representative samples inclusive of different categories in the population.

4.5 Control groups

Control groups are sometimes used in conventional impact assessment to compare communities that have received support with those that have not. In a controlled environment such as a laboratory, this makes sense. But with any real-world impact assessment, use of control groups brings problems of:

- comparability: isolating a control from project impacts
- logistics: to be strictly comparable, the control group should be the same size as the measured group, and
- cost: this doubles the cost.

With resilience-building projects, the problems are often even more acute.

- Is a control group denied access to early warnings to see how those who did get the early warning survive compared to those who did not? Clearly this would be an unacceptable moral hazard.
- Asking communities or households to allocate time to participating as a control group in a process from which they gain no added resilience may erode their resilience.
- Information about a resilience-building measure, such as a seasonal forecast, will flow beyond a specific project area (as indeed it should, to promote scale-up) and is likely to be taken up by a control group who will know about it just by virtue of being a control group. This erodes their value for comparison.
- Logically this means selecting control groups far away enough to avoid this problem, but this is likely to result in a control group in an significantly different context culturally, socio-ecologically, economically, etc., so of little use for comparison.

For these reasons, conventional control groups do not work well with resilience impact assessment. Impact assessment that relies on before-and-after comparisons rather than comparison of participating and non-participating populations, for example, through randomised control trials, is referred to as non-experimental design. A weakness of this approach is the necessity to demonstrate that alternative reasons for impact are either not relevant or not as important as the intervention. This can be resolved by careful attribution and triangulation using more than one assessment methodology across both qualitative and quantitative impacts to verify (or not) the logic of the impact pathway. Both experimental and quasi-experimental design are often held up as the gold standard of assessment, but given the problems associated with control groups (above), non-experimental design measures resilience impact more effectively.

However where valid comparators exist, they should be used, including general data sets for the project area. Often these are limited in extent, may not measure the right things, are sometimes difficult to access or poorly and erratically maintained, but they can be useful for basic measures, for example, average crop yields, so should be applied where comparison can add value to the analysis.

Another option is to compare samples that are different stages in the intervention. So a mature project area that has had several years' involvement in resilience-building activities can be compared with an

Below: Making sure everyone has access to weather forecasts is a key part of climate resilience, but these can easily flow beyond the direct participants in an intervention. Here village noticeboards ensure all can get the right information but are placed near markets so that neighbouring areas also benefit. This renders the use of control groups largely valueless.

area or set of communities that have only recently joined. This avoids the moral hazard problem and ensures that lessons learned in assessment processes used at the beginning of an intervention can then be applied as work progresses, for example, for baselining the recently-joined population. If the recently joined group is assessed early enough, it can also give an almost-complete picture of comparative impact. It does, however, require a project that is designed to add participants in phases rather than working with one set of communities from the outset. The same type of approach can be used where a project has been designed to work at different levels of intensity with different groups.

With shocks, the experience of previous events of similar severity can also be useful. When Cyclone Phailin struck the east coast of India in 2013, it did so at similarly intensity and location to Cyclone 05B in 1999 (see indicator 17 above). Only 45 people died as a result of Phailin, compared with 9,800 for 05B, demonstrating an impressive 99.5% mortality reduction impact of early warning and evacuation to safe shelter.

Below: Farmers in Tanzania explain the design and results of their ongoing on-farm crop research into drought resilience. With the right support, communities are increasingly involved in designing their own research programmes to measure their resilience and address their resilience priorities.

4.6 Bias and error

Both quantitative and qualitative impact assessments can be distorted by biases and errors – typically bias tends to affect qualitative impact more acutely, as data error tends to affect quantitative impact assessment (or where qualitative data is transformed into numbers) but it is important to be aware of the susceptibility of whatever assessment methods chosen to both, as one often leads to the other.

a) Biases can affect the assessor (in the way an assessment is designed and implemented) and the informant (in the way information is provided). Attempts to categorise the various biases to which any form of assessment or research is vulnerable typically show 50-60 different types. Some of the more usual biases that occur in the design and implementation of resilience impact assessment include:

- **design bias** – failure to apply the most appropriate design and sampling approach, resulting in an unrepresentative sample that will either over- or under-estimate impact and/or fail to incorporate the needs of participants and data users in assessment design [26]
- **confirmation bias** – the tendency to search for or interpret information in a way that confirms the assessor's preconceptions, for example, from previous experience, ideology, anecdotal evidence of impact or from unrepresentative feedback
- **tarmac bias** – the tendency to sample informants that are close to accessible transport routes, often to reduce the time and cost of data gathering or reduce logistical challenges. The same kind of bias can be applied to other factors, for example, a survey relying on mobile phone owners for an ICT-style response will miss those too poor to buy a phone and therefore quite possibly more vulnerable as a result

‘It is doubtless impossible to approach any human problem with a mind free from bias’

Simone De Beauvoir

- **group attribution bias** – representing the characteristics of an individual group member as reflective of the group as a whole, or the tendency to assume that group decision outcomes reflect the preferences of group members
- **gender bias** – failure to ensure equal consultation across men and women in the assessment process. As risk can affect men and women differently, this may result in certain categories of impact being under-estimated and others being over-estimated.

Biases that affect the answers given to an assessment process, especially when these are descriptive, open-ended and qualitative responses, include:

- **response bias** – a tendency of a respondent to tailor an answer in order to increase the likelihood of further support. This can work both ways – a respondent may deliberately underestimate impact to justify more support, or may overestimate it if the perception is that a successful intervention is more likely to continue [27]
- **seasonal bias** – assessments of impact changing simply due to the timing of the assessment. For example, immediately after a relatively good harvest may result in responses that describe higher resilience impact than the same questions six months later in the middle of the dry season
- **biases of recollection** – there are a variety of cognitive biases that can distort any assessment based on respondents’ memory of past impact. The **recency effect** means that people remember more recent events more easily than more distant ones, which could lead to recent impact being prioritised whether or not it is more significant. In contrast, **rosy retrospection**, a tendency to judge past events more positively than they were actually judged at the time, could lead to past impact being over-estimated and present impact or progress being under-estimated. The **availability heuristic** leads to respondents estimating what is more readily remembered, which tends to be the more unusual or extreme events, could mean that less dramatic, incremental impact related to resilience to stresses is missed in favour of resilience to shocks
- **biases of attribution** – also known as social biases, these include **fundamental attribution bias** where personality-based explanations of impact are over-emphasised and the effect of other, external factors are under-estimated. This could result in resilience being over-ascribed to individual characteristics and the impact of an intervention or technology being under-estimated. Group attribution bias (see above) also affects respondents, particularly when key informants are relied on to represent a group or community.

b) Errors in statistical data range in their causation from poor survey formulation (using the wrong survey tools, inappropriately deriving statistics from qualitative data, failing to select a representative sample) and analysis, to outright data manipulation. Common problems include:

- **using loaded questions** – questions that signal a ‘correct’ answer in advance to the respondent can result in over-stating the quantitative (and qualitative) impact of an intervention

- **inappropriate methods** – the wrong measuring tools can consistently over or under-estimate results, e.g. the square cut method to measure crop yield in Africa consistently over-estimated productivity by up to 40% as enumerators tended to select atypical field positions (usually in the centre, thus also illustrating design bias) for the square to ensure they had a crop to measure
- **misunderstanding error** – using a non-random sample and using averaged errors for the whole population when much larger errors occur in sub-groups are just two factors that can lead to statistical error being under-estimated
- **fatigue error** – an understandable tendency that creeps in at the end of a day's surveying or towards the end of the survey, especially with over-large sample sizes. Questions are asked in a more hurried way, answers are accepted at face value without follow-up questions to verify their accuracy, and the process generally becomes more of a 'tick-box' exercise
- **data mining** – selectively reporting results because they are consistent with initial objectives and hypotheses, inappropriately rejecting outliers and failing to correct for missing data all affect conclusions about the impact of resilience-building activities
- **data manipulation** – particularly common where there are conflicts of interest, data manipulation includes selectively presenting data to justify a hypothesis, deliberately omitting data that would undermine a hypothesis and managing a sample or changing a baseline until an assessment generates the 'right' answer. These procedures result in assessment results being rejected when discovered.

The important thing is to acknowledge where these are likely to occur and design sampling, data-gathering tools, and triangulation and attribution measures that, as far as possible, either mitigate the bias and/or error from occurring or minimise its effects.

4.7 Using tablet technology

Traditional data gathering has always required pen and paper. With modern ICT technology, such as smartphones and tablets, this is changing. Surveys can be designed in software programmes, such as Kobo,[28] and then uploaded onto tablets or mobile phones to use for data gathering. With increased availability of mobile phones in remote areas, questionnaires can be designed with the communities involved, run either by enumerators or community members and then uploaded back to a central data processing and analysing point. With the importance of community-defined indicators to measuring resilience, this greater interactivity offers potential solutions to defining more context-relevant measures, as well as promoting greater accountability in the impact assessment process.

The benefits of using this technology include:

- processing the questionnaire on to a tablet-based questionnaire format encourages the survey designer to think through the logic of the questions and make amendments that improve their flow
- in-survey mistakes can be more easily picked up and corrected while the data gathering is in process, thereby reducing the risk

Below: Making new connections - mobile phones increase access to early warning and forecasts, but also offer the prospect of new ways of gathering evidence.

of having to return to sampled responders and repeat running questionnaires after the main process has completed

- progress can also be followed and checked using the GPS function to avoid locational biases creeping into the survey, for example, too many responders clustered in one place
- significant cuts to data processing time and cost as this can be directly uploaded on to a central computing point
- allowing for additional data recording with little extra effort or time, such as photographs (with their agreement) of responders, their houses, etc., and GPS positions of responders that in turn can be used to map the sample and show response patterns by geographical area.

However, there are also limitations, such as:

- it can only be used for two of the 30 tools (listed in Annex 2) – questionnaires and formal interviews – and is only really suitable for questionnaires with closed-ended questions (multiple choice or yes/no answers). Theoretically writing in answers is possible, but if this extends beyond numbers, the existing designs of tablet stylus are not satisfactory for writing extensive descriptive answers (voice-enabling could solve this problem by allowing an enumerator or subject to add detail verbally over the closed-ended responses in the survey software)
- it is basically a replacement for a paper schedule – extreme care needs to be taken to manage any soft or hardware malfunctions and ensuring that charging facilities are available, so that the survey does not grind to a halt if functional problems arise or the tablet runs out of power or enumerators accidentally delete a week's data-gathering by pressing the wrong button when uploading
- data still needs to be analysed and interpreted and by a software-capable analyser
- a risk that the technology seduces impact assessors away from other more suitable methods because of its ability to produce superficially attractive graphics and maps
- the start-up costs in terms of tablet purchase considerably exceed the equivalent paper-based systems (but once running, can be reused extensively, thereby improving cost effectiveness).

Like any data gathering, it still relies on designing a survey that asks the right questions about resilience impact in the right way that is relevant to the population being assessed. So it needs to be used in combination with other tools so that open-ended and/or qualitative data can be used to triangulate – this is no different from paper-based systems.

Below: Farmers in Malawi explain how access to gravity-fed irrigation helps them manage dry spells and increase their resilience to drought. A reliable supply of irrigation water depends on a healthy and resilient catchment forest on Mulanje Mountain, above them.

5. Data analysis, presentation and communication

While most of the focus and energy of impact assessment tends to be expended on designing and managing data collection, once this is complete the assessment is **only at the start of the most important stage** – the processing, analysis, presentation and communication of impact. If these are not completed satisfactorily, all the effort involved in designing, implementing and managing data collection processes will be wasted.

5.1 Processing and analysis

Processing methods depend on the type of data collected – so quantitative data may be presented as tables or more interestingly, graphs and charts. Where larger samples are used, more statistical methods can be deployed, including computer programmes designed to handle data and determine statistical significance, such as SPSS.

Qualitative data requires a descriptive approach, typically illustrated with photographs, videos and case studies. Where ranking or scoring exercises have taken place, qualitative information can be presented in numerical formats. This can improve presentation and facilitate more in-depth statistical analysis, particularly with large data sets, while taking care to acknowledge the inherent biases that can occur with qualitative methods.[29] To address these, however, any analysis should seek to show how quantitative and qualitative data triangulate each other to improve validity of results.

The more data you have, the more complex and time-consuming processing and analysis will be, leading to higher costs and longer delays in finalising the assessment report. This bottleneck can be critical in delaying final products, especially where the data collection process:

- uses excessively high sample sizes, resulting in large numbers of interview schedules to process and analyse
- measures lots of ‘nice to know’ rather than ‘need to know’ variables, unnecessarily increasing the complexity of the analysis process and potentially reducing time spent measuring important impact
- measures variables that are poor indicators of the impact being investigated, making attribution to project activities problematic and difficult to process.

On the other hand, unexpected impact highlighted by project participants can add useful detail and highlight areas for more in-depth assessment in later assessments. These issues demonstrate the need to factor processing, analysis, report write-up and communication into the assessment workplan to ensure the final report gets to the right people, in the right format, at the right time. Key impacts (positive, not occurring, and negative) and the factors contributing to their occurrence (activities undertaken by the intervention, as well as other factors beyond its control) should be summarised in the Conclusions section, with follow-up issues described in the Recommendations section.

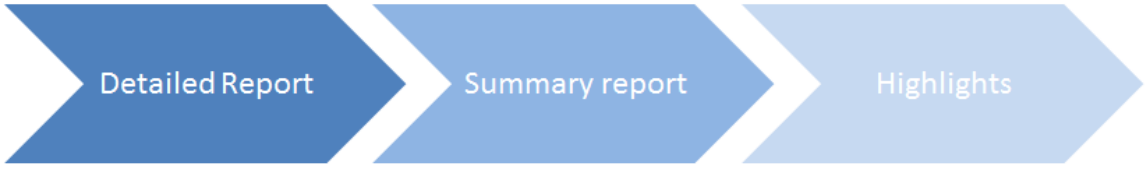
‘The single biggest problem in communication is the illusion that it has taken place’

George Bernard Shaw

5.2 Presentation

A key issue with both presentation and communication is the intended audience for the impact assessment. Typically the results will be required by multiple stakeholders and therefore a range of formats may be required, from a detailed technical report to a multimedia presentation.

Figure 14: Types of report



Detailed Report	Summary report	Highlights
Typical audience • Project managers • Community/community managers • Donors • Technical advisors	Typical audience • Policymakers • Advocacy advisors • Other practitioners • Workshop presentations	Typical audience • Fundraisers • General public • Conferences
Typical Contents • Detailed narrative • Statistical data • Charts and graphs • Case studies • Photographs • Maps • Annexes with working data • Description of methodology	Typical Contents • Summary narrative • Key charts and graphs • Key illustrative photographs • A summary case study or two	Typical Contents • A case study • A headline graph or chart • A key photograph • Video interviews

It is particularly useful to ask at the outset of an impact assessment what sort of evidence and in what format the eventual information users are going to want. For example, if the evidence is to promote the scale up of a particular resilience-building process into policy, it is important to find out what makes a persuasive argument from the perspective of policy makers. If they put an emphasis on quantitative evidence in a highly summarised format, this should be factored into assessment design and delivered in the appropriate format. Programme management, on the other hand, may require more in-depth detail to guide the next stages of intervention planning, so will use a detailed report.

Often those receiving summary or headline results may require a more detailed report later or ask for a reference to it to back up the information they are receiving – this makes relying exclusively on more summary products or unprocessed data for evidence an unwise strategy. So preparing a detailed report is the basic requirement of an impact assessment and from this, other less technical and more accessible products can be generated.

5.3 Communication

A key priority should be ensuring that any external assessments are fed back to the intervention population, usually in a pre-finalisation draft, so that results can be:

- assessed from their perspective and any recommendations as to changes in the intervention theory of change and/or implementation can be made.

- impact measured can be either verified or questioned.
- priorities for further impact assessment can be determined for later investigation.

In the age of social media and global internet connectivity, communicating impact is essential to increase the influence and scale up of resilient development.[30] Inevitably this means tailoring the final product to the communication methods chosen (as per Figure 14 above), and using various options from direct face-to-face briefings and presentation to web-based blogging, tweeting and you-tubing. Detailed guidance on producing effective communications can be found in Your Guide to Producing Great Communications – Christian Aid (June 2014) (see Annex 1, Further Resources).

Below: Going that extra mile to get the message across – using the most appropriate communication method for the audience you are targeting is important, as demonstrated here in the Philippines

6. Conclusions

A number of studies take risk management as an entry point for operationalising and measuring resilience.[31] These provide practical programming tools as they describe the attributes that have been proven to protect lives and livelihoods from shocks and stresses. This bottom-up and experience-based derivation of resilience measures in the context of risk management has advantages, although measures of resilience more broadly have their critics. Concerns about popular measures include:[32]

- their deterministic approaches that focus on inputs, outputs and outcomes rather than processes
- their capture of a static rather than a dynamic picture
- their narrow focus on system effectiveness and efficiency, rather than assessing processes of transformation and resilience.

More experience is needed to compare, contrast and link methods of measuring resilience and risk management effectiveness. The tools developed for impact assessment (see Annex 2) do not need to be re-invented, but measuring resilience impact does have additional challenges. Often increased resilience means a negative impact from a shock does not happen or is significantly reduced. Measuring a negative impact that does not occur is clearly more challenging than measuring a positive impact that does.

While conventional impact assessment equates a lack of progress with failure, resilience impact assessment would focus on whether a livelihood has absorbed a shock or stress, recovered, maintained its functions, and increased its ability to do all of these in the future. So a livelihood experiencing positive resilience impact may still go backwards (at least in the short term), but just not as much as it would have done in an absence of that added resilience.

As we are focusing on resilience of people (as opposed to resilience of materials), the issues of adaptive capacity – our ability to anticipate risk – and thriving so as to enhance future resilience also become important. A resilient livelihood seeks to minimise loses and maximise gains made when short-term shocks are not occurring and beneficial conditions are. So this implies:

- an approach based on non-experimental design and theory-based impact assessment [33]
- a particular priority given to measuring impact over a longer time frame, understanding that more resilient livelihoods demonstrate negative as well as positive impacts
- ex-post assessment to understand how resilience building has continued (or not) after an intervention is completed
- putting capacity to anticipate, manage and recover from short-term hazards as well as mitigate longer-term incremental risk at the heart of the impact assessment process
- balancing qualitative impact (for example, capacity to understand risk forecasts and apply them to resilience-building decisions, develop a PVCA, manage risks and recover fast, etc.) with both financial and non-financial quantitative impact (building assets, savings, increasing yields, returns to marketing, etc.)

‘Keep it simple! This is the cardinal rule for M&E systems. M&E arrangements must be implementable and feasible within the financial and human resources available. Often the more elaborate the system and data requirements, the less likely it is that the system will be able to deliver on its promises’

World Bank

- so that statistical measures are combined with qualitative, contextual information that can triangulate with this data and demonstrate whether or not these statistics are reliable
- and qualitative data is rendered less biased and more representative when backed with statistical measures.

Measuring the impact of resilience building should, of course, use existing tools, but understanding the complexity of enhancing the five key features of resilience (as per Figure 3) will require a range of both conventional and unconventional impact assessment methods in order to obtain evidence that is both valid and reliable. Baselines can be a useful part of this, but ideally should be limited to focusing on a limited and manageable number of key risk factors and community-determined benchmarks that, when re-measured, will reveal impact that can be confidently measured and that demonstrate the contribution of the support given.

In addition to selecting the right resilience indicators for the context being measured, the Overseas Development Institute (ODI) has developed nine key best practices (see Figure 15 below) with respect to measuring resilience to disasters, but which can also be seen as relevant to resilience to stresses.

Finally, it is important that community members and their governance structures are empowered to assess impact and use it in a way that enhances accountability in the management of resilience plans (as per the SPICED in 3.2). If impact across communities can be pooled from this process, this can provide evidence for the essential policy and practice change needed to enhance resilience at local government level, such as ensuring community action plans are formally integrated into local government development plans, which themselves can be aggregated into national plans to be resourced from climate change adaptation funds and other sources. This local action plan to national policy and spending level, if backed with good impact evidence, can increase the chances of development and adaptation funding getting down to the most vulnerable, where it is needed most.

Below: Through participatory photography, groups in Kenya share evidence of how they have built resilience in their communities. This ensures that the impact is measured from their perspective, according to their priorities and using their continual assessment processes. Here group members are explaining to each-other what each set of photos mean, what are the lessons learnt and consequently the implications for further resilience-building.

Figure 15: ODI's nine ways to measure disaster resilience better

ODI has analysed 50 frameworks to distil essential insights on measuring resilience. Based on the findings of these analyses, it recommends the following best practice approaches:

1. **Draw on the key concepts of resilience**: resilience-building approaches draw on a wealth of research and thinking. These should be reflected in approaches to measurement, embedding ideas such as 'systems thinking' within any proposed metrics.

2. **Measure often and remain flexible**: resilience thinking helps us understand complex interactions within a system, and account for continual evolution and a constant state of change. This is why gauging the degree to which a system is resilient requires a higher frequency of measurement and a frequent adjustment of approaches.

3. **Collect disaggregated data:** DRR monitoring and evaluation does not systematically rely on data broken down by gender, age, disability, ethnicity and socio-economic status. Data collection needs to be disaggregated in order to understand the impact of disasters on vulnerable groups and ensure that resources are targeted equitably.

4. **Measure across systems and scales**: measuring the performance of different, inter-related systems across different scales is vital. This includes human and institutional systems, as well as economic, environmental or infrastructure systems. Actions to strengthen all or part of a system need to be measured in the context of the wider systems in which it operates.

5. **Balance assets and processes**: frameworks should track and measure both assets (schools, savings, health infrastructure) as well as processes (good governance, transparency, inclusion) as part of a holistic approach to measuring resilience. Also, it's vital to measure the quantity and quality of assets and processes, therefore 'thresholds' must receive adequate consideration in any approach to accurately gauge resilience.

6. **Include the political context**: what constitutes 'good resilience'? When developing frameworks, it is critically important to ask framing questions, such as 'resilience of whom, by whom and to what ends'. Finding ways to measure and track these over time remains a real challenge.

7. **Track hazards and impacts:** assessing hazard intensity and the impact on livelihoods is necessary to accurately measure the effectiveness of resilience programmes. If impacts remain the same or reduce as hazard intensity goes up, then we can say that a household is more resilient.

8. **Track risk of loss, not just losses:** as major disasters happen infrequently, simply measuring deaths or economic losses over a certain time period may provide misleading results. It is therefore vital to measure the 'risk of loss' from a range of hazards to understand whether resilience-building interventions are making progress (for instance, through the use of models to calculate risk).

9. **Invest in high-resolution data:** high-resolution data is needed to measure risk using the real experiences of disaster losses to validate findings. Improving the availability of such data is a critical challenge, requiring increased and coordinated resources from both public and private sectors.

Annex 1. Further resources

A. Resilience-specific resources

- *Characteristics of a Disaster Resilient Community* – Stephen Twig (DFID, 2009)
- *Learning to Adapt: Monitoring and Evaluation Approaches in Climate Change Adaptation and DRR – Challenges, Gaps and Ways Forward* – Paula Silva Villanueva (SCR Discussion Paper 9, 2011)
- *Towards a Quantifiable Measure of Resilience* – Christophe Béné (IDS, 2013)
- *Toolkit for Measuring Community Disaster Resilience* – GOAL (2014)
- *Disaster Resilience Measurements: stocktaking of ongoing efforts in developing systems for measuring resilience* – UNDP (2014)
- *Measuring and Assessing Resilience: broadening understanding through multiple disciplinary perspectives* – Quinlan et al (Journal of Applied Ecology, 2015)
- *Measuring 'Subjective Resilience' using People's Perceptions to Quantify Household Resilience* – Lindsey Jones and Thomas Tanner (ODI WP 423, 2015)
- *The 3As: Tracking Resilience* – Bahadur et al (BRACED, 2015)

B. General guidelines and resources

- *Planning for Participatory Monitoring and Evaluatio*n – John Plastow (SOS Sahel, 2000)
- *Monitoring and Evaluating Advocacy* – Jennifer Chapman and Amboko Wameyo (Jan 2001)
- *NGOs and Impact Assessment* – INTRAC NGO Policy Briefing (2001)
- *Methods Manual for Fieldwork* – The LADDER Research Team (LADDER Working Paper No.2, September 2001)
- *Practical Guidelines for the Monitoring and Evaluation of Capacity Building: experiences from Africa* – Rick James (INTRAC, Oct 2001)
- *Outcome Mapping: building learning and reflection into development programs* – Sarah Earl, Fred Carden and Terry Smutylo (IDRC, 2001)
- *Planning, Implementing and Evaluating Capacity Development* – Douglas Horton (ISNAR, July 2002)
- *Impact Monitoring and Assessment* – Karl Herweg and Kurt Steiner (GTZ, 2002)
- *Ethical Guidelines* – Social Research Association (2003)
- *Participation and Numbers* – Robert Chambers (PLA Notes, 2003)
- *A Typology of Sampling Designs* – Anthony Onwuegbuzie and Kathleen Collins (The Qualitative Report, June 2007)
- *Methodological Guide for Evaluation of Pro-Poor Impact of Small-Scale Agricultural Projects* – Paz et al. Centre for Development and Poverty Reduction (Imperial College, 2006)
- *Managing for Impact (M4I)* - Cecile Kusters and Clare McGregor (Wageningen, 2010)
- *Time to Listen: hearing people on the receiving end of international aid* – Mary B Anderson, Dayna Brown and Isabella Jean (CDA Collaborative Learning Projects, 2012)
- *Participatory Impact Assessment: A Guide for Practitioners* – Andrew Catley, John Burns, Dawit Abebe and Omeni Suji (Feinstein International Center, Tufts University, 2013)
- *Impact Evaluation: A Guide for Commissioners and Managers* – BOND (2015)
- *Evaluation Policy for Christian Aid International Programmes* – Christian Aid (2015)

For Christian Aid staff, copies of resilience-specific resources, general guidelines and resources and specific tools (as per Annex 2 below) can be found on the intranet on this link:

Measuring Resilience Impact Resources

New resources will be added to this as they are produced.

Annex 2. Methods and tools for collecting information

Method	Overall purpose	Advantages	Challenges
Documentation review (secondary data)	To gain an impression of how an activity operates without interrupting it, from review of project records and other relevant information – financial records, surveys, project proposals, stakeholder records, background situation analysis, etc.	- Gets comprehensive or broader historical information. - Doesn't interrupt activity or partner's routine in activity. - Draws on existing information from other sources. - Few biases about identifying information.	- Often takes a great deal of time. - Information may be incomplete. - Need to focus on the activity and avoid diversion. - Not a flexible means to get data; data restricted to what already exists.
Observation	To gather accurate information about how an activity or process actually operates, practically.	- View operations of an activity or process as they are actually occurring. - Can adapt to events as they occur.	- Can be difficult to interpret observed behaviours. - Can be complex to categorise observations. - Can influence behaviours of project participants. - Can be expensive or time-consuming.

Mainly **quantitative:**

Method	Overall purpose	Advantages	Challenges
Questionnaire survey, checklist	To extensively and comprehensively obtain a lot of information from people.	- Can be completed anonymously. - Inexpensive to administer. - Easy to compare and analyse. - Can be administered to a large sample, getting a lot of data. - Sample questionnaires usually exist for guidance. - Can be applied using tablet technology, reducing data processing time.	- Might not get careful feedback (avoiding the subject telling you what you want to hear). - Wording can bias responder's responses, might be unclear locally. - Are impersonal, can be alienating, threatening and time consuming for responders. - Need sampling expertise. - Do not get full story, limited use for revealing complexity, will not reveal unanticipated impact. - Matching data gathered with capacity to process into a useful report.
Formal interview	To fully understand someone's impressions or experiences, or learn more about their answers to questionnaires.	- Get full range and depth of information. - Develops relationship with the respondent. - Can be flexible with the respondent. - Can be applied using tablet technology, reducing data	- Can take a lot of time. - Can be hard to analyse and compare. - Can be costly. - Interviewer can bias direction of investigation and client's responses.

		processing time.	
Statistical assessment	Post-production measurement of seasonal or annual output	- Hard data on productivity changes that can be used in gross-margin and cost-benefit analysis. - Hard data from existing experience.	- Methodologies subject to unreliability. - Time and expense in achieving statistical significance and processing data.
Community mapping (geographical application)	Provide geographic layout of community area for study and subsequent tool to overlay data on a range of aspects.	- May produce valued first visual record for community. - Highlights different perceptions of physical or social environment. - Can be triangulated with GIS maps.	- Transferring accurately from ground to paper. - Interpretation of key features by outsiders, without community to advise. - Subjectivity in recording local features.
Transect walk	Get a 'snapshot' (geographical record) of community area by walking through it.	- Covers main ecological, production and socially stratified zones of a project area/community. - Direct contact with people and the environment. - Identifying impact collectively.	- Extracting maximum information from local 'experts' e.g. community leaders. - Understanding context of data.
Wealth/ vulnerability ranking	To rapidly stratify a community into perceived wealth, vulnerability or social groups.	- Gives a set of criteria that in the eyes of the local population indicates what constitutes wealth/vulnerable groupings. - Combines qualitative and quantitative indicators. - Gets around the sometimes controversial process of measuring direct income.	- Community representatives may sanitise differences that characterise different wealth/vulnerable groups. - The results can be at odds with conventional methods of classifying people, e.g. solely according to income. - Gender differences may be hard to distinguish from household differences.
Gross margin analysis	To understand whether the returns to a particular enterprise exceed the variable costs of implementing it.	- Provides short-term information on the profitability of an enterprise. - Provides a clear quantified assessment that can be used for enterprise planning. - Can assist enterprise operator to identify reasons for low gross margin and make key adjustments.	- Labour costs are notoriously difficult to assess and value (so often presented without). - Factor (inputs and outputs) prices may fluctuate and undermine use of GM data to plan future enterprises. - Need to consider other factors, e.g. risk, market access, environmental impact.
Financial self-sufficiency index	To determine the income and expenditure of an organisation and the path to financial sustainability.	- Enables quantification of income streams and identification of sustainability point. - Identifies priorities to be addressed within the project period by both the institution and capacity-building partner, sharing responsibility and	- Budgeting skills to quantify accurately income and expenditure. - Income may fluctuate, be unpredictable, requiring frequent revision of index.

		promoting local ownership. - Can stimulate the institution to take important measures and overcome key hurdles.	
Cost-benefit analysis	To quantify existing costs and benefits, and project likely future costs and benefits to determine the net present value of an intervention.	- Balances costs against benefits, giving a forecast of likely future returns. - Widely used by the public sector, so can be persuasive for 'value for money' evidence and advocacy. - Potential for finance staff to get involved in impact assessment.	- Identifying all relevant costs and benefits. - Difficulties arise when costing qualitative change. - Discounts the future, particularly when future externalities are difficult to or not estimated. - Complex processing and interpretation capacity needed to avoid misuse.

Mainly **qualitative:**

Method	Overall purpose	Advantages	Challenges
Focus groups	To explore a topic in depth through group discussion, such as reactions to an experience or activity, understanding common constraints, etc.	- Used quickly and reliably get common impressions. - Efficient way to get a range and depth of information in short time. - Can convey key information about projects. - Combines well with other tools, e.g. spider diagrams. - Can use polling techniques and aggregate to increase representativeness of data.	- Can be hard to analyse responses. - Need good facilitator to avoid domination by powerful voices. - Works best with relatively small groups (up to 20), so can be difficult to schedule optimal - sized groups across a project.
Case studies	To fully understand or depict a subject's experiences in a project, and also to conduct comprehensive examination through cross comparison of several cases.	- Fully depicts subject's experience in project input, process and results. - Powerful means to portray project to outsiders. - Good triangulation tool for other, more quantitative methods.	- Usually quite time consuming to collect, organise and describe. - Represents depth of information rather than breadth, so tendency to bias if unrepresentative subjects or star performers are chosen.
Most significant change (MSC)	To use representative change stories to assess complex participatory processes, usually through several layers of selection processes to refine the stories or case studies into an unbiased picture of impact.	- Involves all stakeholders in selecting representative change stories. - Reveals complexity and unanticipated impact. - Can significantly reduce the bias of individual case studies. - Easy to apply with existing case study collection skills. - Does not need highly technical processing of information.	- Can be subject to bias in the selection process. - Needs to be set up at the start of projects to ensure continuous collection of change stories. - Defining what a significant change is may be different in different contexts, so making comparison difficult. - Can be seen as competitive and time consuming.

Gender analysis frameworks	A range of tools designed to understand the differential impact of an activity on women and men.	- Understanding both practical and strategic impact on women and men and how it changes the relationship between them. - Different frameworks can be applied depending on the context and difficulty of the issue, given the existing cultural context. - Enables a more representative rather than partial (and usually male-centric) focus on impact.	- Facilitation skills needs to ensure that representative impact can be determined from women and men. - Cultural contexts can make using the tools difficult. - Need to ensure a gender balance in the impact assessment team (which is a relevant challenge for all tools, but especially important here).
Strengths, weaknesses, opportunities and threats (SWOT) analysis	Records the key issues relating to the internal and external environment, and the relationships between them.	- Enables communities to identify key issues. - Promotes consideration of the wider environment. - Can unblock key hurdles to progress.	- To ensure all issues are exposed, even from more marginalised/poorer community members. - Exposing weaknesses can be difficult.
Beneficiary/ participant assessment	Involves the participation of beneficiaries in evaluating a planned or on-going development activity and builds on the experience of participant observation.	- Involves participants by providing a context for quantitative data by letting their voices, values and beliefs be expressed.	- Needs careful facilitation, especially where assessing internal/difficult to expose risks. - More confident community members' voices and values can dominate.
Historic profile (or timeline)	To summarise in time, the main historic community events and trends within living memory.	- Provides an opportunity to involve and draw on memories and experiences of all members of the community, especially the oldest. - Human, personal or community view of history. - Triangulation with background data to confirm or correct remembered events/trends and add context and detail to data.	- Drawing out history from the oldest community members and reasons for changes or trends. - Memory can be unreliable.
Testimonial (portrait)	To record a person's thoughts, feelings and experiences.	- Carried out in the first person narrative style. - Learning about a project or its impact through the voices of participants and stakeholders. - Help to reveal the degree of empowerment. - Help to corroborate other sources of data and information and provide a more personal insight into a project's achievements.	- Reality versus expectation. - Personal not communal, so may not be representative. - Relevant to focused rather than incidental detail. - Sensitivity to potentially fruitful new areas of investigation.
Force field analysis	Used for listing, discussing and evaluating the forces for and against a proposed change.	- Determines if a proposed change can get support. - Identifies obstacles to successful solutions. - Suggests actions to reduce the	- Subjectivity in determining the relative importance of both driving and restraining forces. - Imperfect knowledge of forces more distant to the proposed change.

		strength of obstacles.	
Matrices (Pair Wise Preference and Direct)	To identify and prioritise impact (e.g. basic needs/problems/ issues/solutions and reasons).	- Rapid assessment providing much detail. - Powerful comparative and investigative tools. - Can be used across a wide selection of contexts (including impact assessment of peace building-related activities). - Can illustrate qualitative change with numbers.	- Ensuring real needs (framing problems) are identified and for various social groups and gender separately. - Process to action, can raise expectations if no 'follow-through'.
Seasonal calendar	Use visual representation to help to identify events that occur seasonally.	- Useful for a wide range of projects and activities (planting, harvesting, identifying seasonal transmission of malaria and other health risks, etc.). - Highlights timing of constraints, such as drinking water availability, labour availability, food intake, illness, etc.	- Translating a seasonal calendar created on the ground using natural objects, e.g. stones, sticks, or leaves to paper. - Identifying issues that relate to extreme seasons (which may not have been recently experienced).
Semi-structured interview	Informal interview with minimal agenda.	- Allows for the development of conversation and the reciprocal transmission of information by subject. - Subject is participating in the discussion (rather than just responding), allowing open-ended and follow-up questions to be discussed.	- Allow the subject to direct the interview. - Need good skills to carry out interview, e.g. not to lead the subject's responses. - Processing of diverse responses can be difficult; subjectivity can creep in giving undue weight to some information at the expense of others.
M&E wheel (or spider diagram)	Allows community-based organisations and their members to self-assess progress towards agreed objectives.	- When used at regular intervals, gives a useful picture of progress in capacity-building, etc. - If indicators are ranked, highlights priority issues for next stages of project. - Can illustrate qualitative change with numbers.	- Works best with no more than five or six indicators. - Ensuring agreement and involvement of all group members on setting indicators and assessing progress. - Subjectivity in self-assessment of situation before and after. - Aggregation can result in bias towards the mean.
Venn diagram	Look at relationship and status of all community and related external organisations.	- Illustrates different participant perceptions of access to and value of resources or of social restrictions.	- Encourage communal assessment of local organisations and groups and relationship to related outside organisations.
Mapping (social application)	Provide tools for establishing the social interactions (or lack thereof) within the community, the resources that are available and accessed to/by them.	- Highlights inter-relationships and inter-dependencies in a group or the community. - Can be used at various stages of a project.	- Important to ensure a good cross section of participants, including non-participants.

		- Focus on contribution rather than attribution.	
Participatory video (PV)	A set of techniques to involve a group or community in shaping and creating their own film.	- Allows communities to select priority issues to film. - Shows a visual representation through film and the direct voice of the community. - Creates confidence and motivation on the part of community members.	- Can sideline less confident members of the community or those uncomfortable with being portrayed visually for cultural or other reasons. - Editing to representative footage without bias. - Technical knowledge and equipment required. - Size of files makes integrating into a report difficult.
Photographs/ participatory photography	To allow visual representation of activities and changes.	- Can greatly enhance narrative information. - Adds a new dimension to reports, makes them much more interesting to read. - Useful for triangulation and ground-truthing	- Need to be used in context, as may not show the whole story. - Should always obtain subject's permission before use. - May be culturally inappropriate in some circumstances.
Drama and song	To reveal community reactions to programme (or existing situation/history).	- Uses traditional cultural methods of communication. - Involves community in public appreciation, response and display.	- Encourages emergence of deep feelings and priorities in community which need to be handled carefully.

Endnotes

1 Both participant and beneficiary are used to mean those benefitting from an intervention, the former being more appropriate for resilience-building. It refers to the households and communities that both participate in and manage project activities and interventions that are facilitated by our partners.

2 *Searching For Impact And Methods: NGO Evaluation Synthesis Study* - Report Prepared for The OECD/DAC Expert Group On Evaluation (1997) and *Methods of Impact Assessment Research Programme, Resource Pack and Discussion* – F. Blankenburg (Oxfam UK/I and Novib, 1995) respectively.

3 Based on information from "A Guide for Project M&E" - IFAD

4 *Partnership for Change*, Christian Aid's Strategy from 2012 states under the principle 'clear thinking' that 'our actions are based on rigorous research and analysis.'

5 *Resilience and Sustainable Development* - Folke et al. (The Environmental Advisory Council to the Swedish Government for the WSSD in Johannesburg, 2002).

6 In more fashionable M&E circles, this continuous process is sometimes referred to as 'evaluative monitoring'.

7 Asset pentagram adapted from the Sustainable Livelihoods Approach (DFID, 2001).

8 In this sense, catchment forests provide multiple parallel buffers.

9 This raises the interesting issue of known and unknown standby buffers – given the relatively limited knowledge of global biodiversity, a substantial amount of potential standby buffer capacity in natural resources is therefore unknown and likely to remain so, given the accelerating loss of species caused by human-driven environmental degradation.

10 Ostrom et al (in *Redundancy and Diversity: Do They Influence Optimal Management?* 2003) conclude with respect to the 'redundancy is inefficient' assertion that *'it is short-sighted and ineffective to derive policy from untested assumptions'*.

11 *Reducing Risk and Driving Business Value*, CDP Supply Chain Report (Accenture, 2012-13).

12 This is also plays an important role in ensuring that standby buffers do not degrade in between shocks.

13 Whether the effect is exclusively or only partly due to an intervention is explored in section 4.1.

14 For further discussion see *DFID's Approach to Delivering Impact – UK Government Independent Commission on Aid Impact*, ICAI, June 2015.

15 See *Monitoring & Evaluation* - BOND Guidance Notes No. 4.3, 2005.

16 Malawi risk assessment developed by communities in the Enhancing Community Resilience Programme, ECRP; Burkina Faso risk assessment developed by communities in the Building Disaster Resilient Communities programme, BDRC.

17 Proxy indicators need clear and objective evidence of the correlation between the impact being measured and the proxy measure, and a clear justification that using a proxy delivers better evidence in a more cost-effective way.

18 This would include the participants as well as the sources of technical advice on risk, such as climate scientists, hydrologists and catchment management experts.

19 Based on actual cases. As risk affects women and men in different ways, it is important to differentiate the impact of resilience-building measures by gender and other relevant characteristics, such as disability and age. See also section 2.6.

20 Ibid.

21 This is a combination of qualitative and quantitative impacts – in practice they often come together (as per indicators 18 and 19).

22 See the *Road to Resilience* - http://resilientlivelihoods.christianaid.org.uk/ for more information, toolkits and guidelines on various aspects of resilience.

23 Frustration with conventional impact assessment's preoccupation with proving attribution has led to the **outcome mapping** approach to M&E planning – further guidance on this can be found through the resources link (under General guidelines in Annex 1).

24 A 95% confidence level with a +/- 5% margin of error shows that impact measured is 95% likely to be in the range 90-100% of actual impact, not simply a 95% certainty of impact. A range of online sample calculators are available e.g. http://fluidsurveys.com/survey-sample-size-calculator/ that can guide sample size based on just confidence level and margin of error, although as the PARCS approach suggests, this should not be the sole consideration.

25 A more detailed discussion on sampling and social research methods can be found in *"A Typology of Sampling Designs"* (see Annex 1, section B.)

26 Conflict of interest is also a form of design bias, common in corporate research where the results are compromised by the researchers' interest in maximising sales of the product being investigated (see also data manipulation in note 25 below). Sampling failures are also referred to as selection bias.

27 Response biases are a wide family of biases that affect respondents' ability to give accurate information – from something as innocuous as a liking for the enumerator leading to answers designed to please them to the more manipulative example (as above).

28 See http://www.kobotoolbox.org/

29 See section 4.6.

30 See also *Christian Aid's Guide for Advocacy in Resilient Livelihoods Programming*, August 2012.

31 See *Characteristics of a Disaster Resilient Community*, Stephen Twig, DFID, 2009.

32 Learning to Adapt: Monitoring and Evaluation Approaches in Climate Change Adaptation and DRR – Challenges, Gaps and Ways Forward – Paula Silva Villanueva (SCR Discussion Paper 9, 2011)

33 This is also consistent with a programme planning approach based on the development of *theories of change.*