# CO IMPACT MEASUREMENT GUIDANCE

***Draft November 2010***

**CARE-USA Program Impact (π) Team**
***All comments and suggestions most welcome!***
***Please send comments and suggestions to vpetrova@care.org***

CARE USA's Country Office Impact Measurement guidance (CO IM guidance) aims to build a common understanding of key concepts, principles and standards of impact measurement (IM) for improved program quality and impact. The FY10 draft of this guidance focuses more on conceptual clarity around the key processes, components and approaches that an impact measurement system needs to support. Throughout the document, relatively more emphasis is given on the need for critical thinking, analysis and reflection than on the technical aspects of measurement, such as indicator selection or data collection. In the FY11 continuation of this guidance, the Pi team plans to introduce operational guidance around methodologies, measurement, and indicator work.

## I. INTRODUCTION

**Why impact measurement?**
With CARE's global commitment to shift to the program approach (PA) and achieve long-term sustainable impact on the lives of the extremely poor, most marginalized and vulnerable population groups, it becomes crucial to enhance CARE's ability to measure and demonstrate impact at CO as well as global level. Measuring impact will help CARE determine how what we do contributes to creating long-term, sustainable change in the lives of our impact groups. Internally, impact measurement ensures that we hold ourselves to the standards of program quality and achieve the desired results. Externally, impact measurement data will be useful to technical, policy advocacy, fundraising, and marketing and communications staff in their work with various stakeholders and constituencies. Finally, impact data will provide the basis for more strategic decision-making and better investment strategies by CARE-USA and other CI members. It will help us answer the question, what difference does CARE make, and what makes us distinctive?

**Why measure impact at the global level?**
CARE is drawing upon global impact measurement, reporting, and analysis as organizational priorities to better tell its global impact story. CARE will measure impact at the global level to improve:

1. ***Accountability to our stakeholders:*** The IM system will help improve accountability to those we work for (impact groups), those we work with (partners, alliances, networks), and those who support and finance our efforts (donors) by generating evidence of long-term sustainable positive impact in the lives of extremely poor, most marginalized and vulnerable population groups.
2. ***Program quality:*** CARE recognizes that our ability to truly evaluate our work and improve our quality depends on our ability to measure impact. IM system will play a vital role in generating knowledge and promote learning rather than just simple measurement. Knowledge and learning generated by IM system will in turn make CARE more effective to address the underlying causes of poverty and marginalization. Being a learning organization, it is not enough for CARE to simply explain what we have achieved; we need to be able to say how and why we have or have not

achieved our desired impact. We also need to know which of our interventions, strategies, and approaches are effective and which need improvement. IM system will help us to continuously learn and improve our program quality by producing timely knowledge, information and evidence of impact.

**Why measure impact at the CO level?**
Another important level at which we need to measure impact is the CO.

1. ***Demonstrating impact:*** The IM system will enable COs to demonstrate accumulated impact at local and national levels beyond project boundary and timeframe, as well as the CO's contribution to relevant MDG and PRS changes that would be measured by global impact indicators (MDI+ indicators).
2. ***Improving program quality:*** Internally, COs will be able to improve their program quality by incorporating knowledge generated through testing Theories of Change (TOCs), continuously analyzing trends and contextual factors, and reflection and learning. All of these capabilities are part of the IM system.
3. ***Influencing policies and development practices:*** The IM system will help COs conduct evidence-based advocacy to influence public policies, and to develop practices that more effectively address the underlying causes of poverty. It will also help enable COs to raise funds for long-term programming by demonstrating significant and sustainable impacts of their work.
4. ***Improving 360-Degree accountability:*** Like global impact measurement, CO level impact measurement provides accountability to those we work for, to those we work with, and to those who support and finance our efforts. The IM system will provide accountability to those we work for by demonstrating long-term sustainable positive impact in the lives of the most marginalized and vulnerable populations. The impact measurement would also support accountability to those we work with and partner with. In addition, this approach will enable donors to see how their investments have resulted in sustainable change.

**Impact Measurement and the Program Approach**

The program approach – our commitment to working for long-lasting sustainable change and measuring our success by the improvement in the lives of impact groups – both necessitates and makes possible impact measurement. There are direct implications of adopting the program approach for what our impact measurement should look like. Below, we spell out those implications with direct reference to what the characteristics of a program dictate for impact measurement.

What does impact measurement look like under the program approach?

| ***Characteristics of a Program*** | ***Implications for IM*** |
|---|---|
| 1) Long-term impact goal for a population at broad scale | The focus shifts away from what CARE is doing, to what changes take place in the lives of the impact group. In traditional M&E, we may have done an evaluation with no measurement or involvement of the impact group. In IM, our focus is on the impact group.<br><br>The broad scale means we are measuring outcomes and impacts for an entire population group (our impact group), rather than just for those directly participating in a CARE project. In IM, we need to look at measurement that covers this entire group, nationally. |
| 2) Thorough analysis of underlying causes of poverty, gender | The analysis of underlying causes of poverty, gender inequality and social injustice is part of the continuous contextual analysis that good |

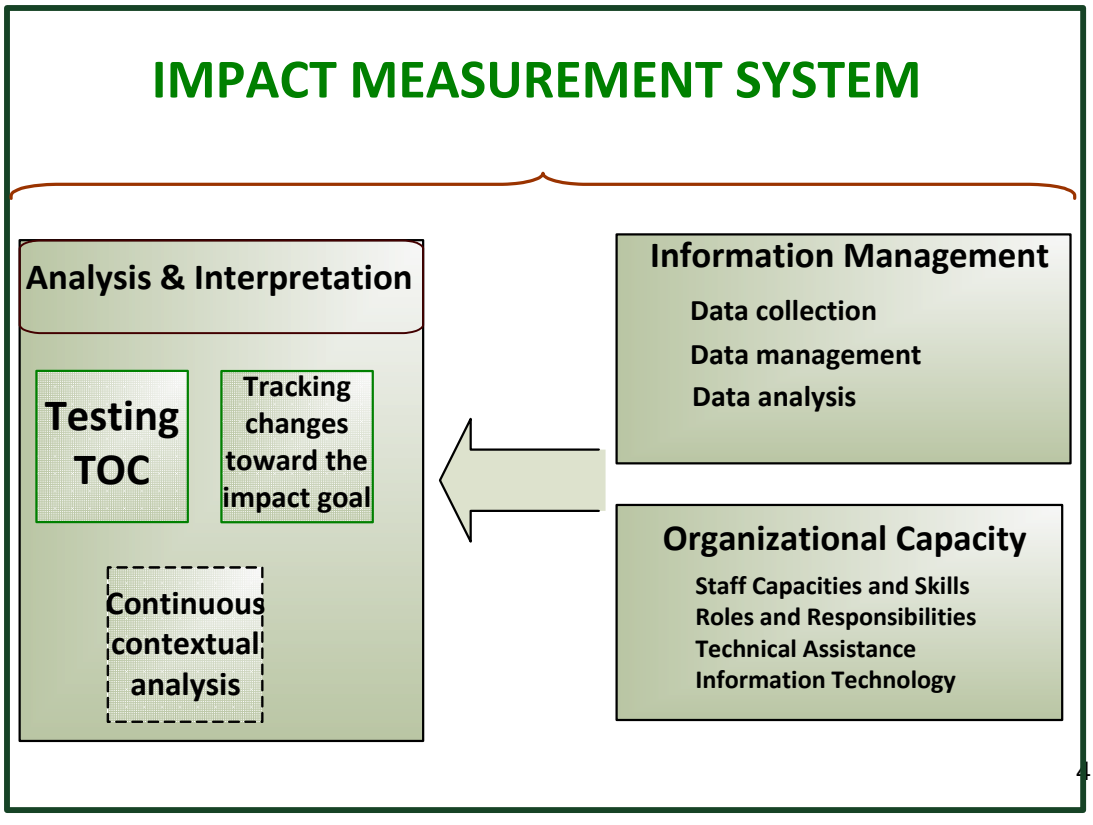
| | |
|---|---|
| inequality and social injustice | programming involves.<br><br>This has two implications for IM. First, we need to incorporate contextual analysis in our impact analysis, to reflect that CARE does not work "in vacuum," but is part of and seeks to change complex social realities. Second, we need to measure and analyze work on the underlying causes of poverty, inequality and injustice. This means measuring on outcome and impact level, not just concerning ourselves with process and outputs. It also means doing analysis to understand and show how outcomes related to social positions and enabling environments impact human conditions. |
| 3) Explicit, rigorously tested and adapted theory of change (TOC) | Testing the TOC is the core process an IM system performs. Since the TOC is central to a program, testing it is central to the IM system of that program.<br><br>For building an IM system, this means: starting with the developed TOC; making its assumptions and hypotheses explicit so that assumptions are laid out and hypotheses can be tested; selecting relevant indicators that reflect the most important social changes the TOC identifies; selecting appropriate methods and applying them rigorously to test the strategic hypotheses; analyzing and drawing implications for how to refine our TOC. IM that does not support testing the TOC does not do its main job. |
| 4) A coherent set of initiatives | In the same way in which a coherent set of initiatives works together toward achieving the impact goal, the measurement and analysis done under each of them contributes to measuring, analyzing and testing the program's TOC and results. We need to coordinate the set of initiatives under each program so that their measurement, analysis and learning come together to form those of the program. |
| 5) Ability to promote organizational and social learning, to generate knowledge and evidence of impact | Our IM systems need to support two functions: the production of evidence of impact and learning based on that. As impact on social change – and our contribution to it – is a complex phenomenon not easily captured in measures, IM systems move away from measurement only and encourage analysis, interpretation, and learning. They also produce evidence of impact using rigorously applied methods so that CARE is able to leverage this evidence.<br><br>Further, IM feeds into our larger learning and knowledge management (L&KM) agenda, rather than the other way around. That is, all IM is done in the larger context and agenda of learning; IM helps move L&KM forward, and contributes to it, rather than knowledge management being done for the sake of IM. If we do rigorous IM but do not use it to advance the learning agenda, we have not done our job. |
| 6) Contribution to broad movements for social change, through partnerships and alliances | Working with others also means measuring and analyzing with others. Unlike traditional M&E, where we mostly focus on our own work and what we can attribute to CARE, IM requires that we look beyond ourselves for the measurement and analysis that would help us understand impact and social change.<br><br>This has implications for how we conceive of measurement and analysis for IM: we need to focus IM on the impact results we are striving for, and share the analytical thinking and learning with our partners. Our definition of partnerships may also involve partnering with actors and agents who share our learning agenda and our measurement and analysis goals. This has practical implications as well: we will often need to use others' metrics and data as much as we can, as CARE cannot measure on national level. |

| | |
|---|---|
| 7) Strategy to leverage financial and other resources | We include the rigorously produced evidence of our impact into our strategies for leveraging financial resources, building partnerships and alliances, and advocating for change. |
| 8) Accountability systems to internal and external stakeholders that are transparent | Accountability, achieved partly by being transparent, is the main reason for doing IM. IM that does not support our being accountable is not worth the time, money, and work we would spend doing it. While it may fulfill other needs as well (such as helping raise resources, do marketing, etc.), its underlying reason for existence should be accountability.<br><br>At the same time, simply doing IM does not mean being accountable. This is a larger organizational goal that cannot be met simply by doing measurement and analysis. So IM is in service of accountability, it is done to support accountability, but being accountable to internal and external stakeholders is a mandate for the entire program approach and for all of CARE. |

**What does an IM system do?**

In order to measure and assess impact systematically and rigorously, we need to build an impact measurement system. The CO-level IM system is mainly designed around CO programs, with the explicit objective to measure and demonstrate accumulated impact of projects and non-project initiatives at the CO level beyond project level. The system is designed to:

- Measure and demonstrate accumulated impact of both project and non-project initiatives
- Validate the program TOC
- Generate knowledge
- Promote learning
- Link to UBORA and the Program & Project Directory to feed information into global systems
- Learn from the analysis and knowledge generated by global systems
- Put learning into practice to improve program quality and achieve greater impact



**What is the relationship of the IM system to other global systems and processes?**

With CARE's global commitment to long-term programming, CARE is going through a profound organizational change process. The transition results into the development of new initiatives and systems as well as modifying existing systems, practices and policies. However, none of these systems, policies or practices is meant to be stand alone. Rather, most of these systems are interlinked and play a complementary role to each other. Similarly, the IM system is directly linked with other systems and process, such as UBORA, the Program & Project Directory, the Strategic Impact Inquiry (SII) process, the PQAT for program design and implementation, and others.

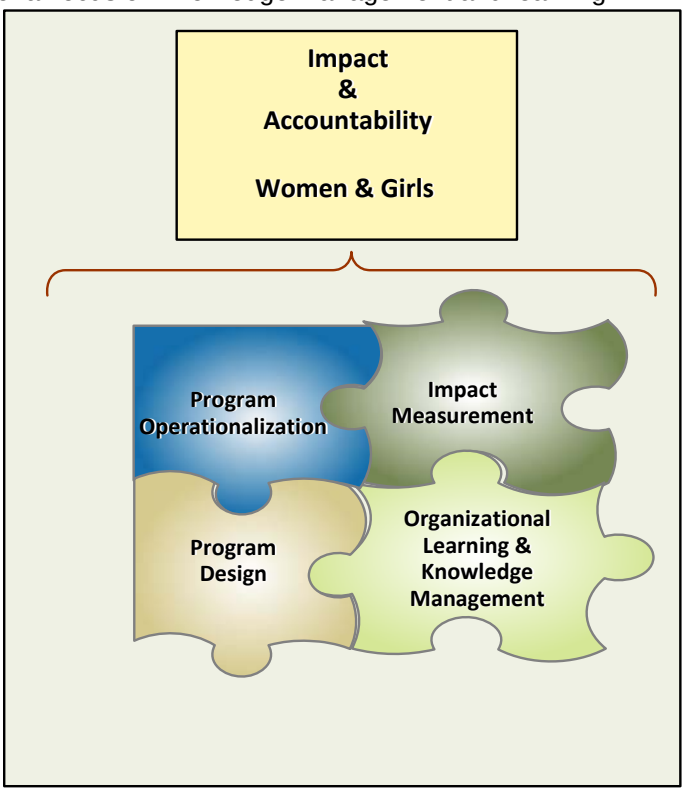
***Impact Measurement and the SII***

[insert text from Maliha on IM systems and SII – PENDING]

***Impact Measurement and Learning and Knowledge Management (L&KM)***

Knowledge management and learning are essential elements of any organization's work, and that is also the case with CARE. All the work that we do as individuals, units and teams in the organization feeds into a collective organizational knowledge base that enables us to carry out our mission. Some of this knowledge is generated formally with explicit analysis, management and dissemination. However, most of this knowledge remains informal and implicit knowledge that resides in institutional memories and scattered in more formal pockets throughout the organization. There is a general acknowledgement in the organization that there needs to be more intentional focus on knowledge management and learning in CARE in order for us to optimize our impact.

The pockets of formal information and knowledge management within CARE consist of systems such as UBORA, PQDL, and LAC's OP&L systems. The IM system will be one of these and will show how the impact of our work in Country Offices has contributed to meaningful change in the lives of the people we serve. As such, our Impact Measurement guidelines are not intended to provide Country Offices, or any part of our organization, with "how to" guidance on Knowledge Management and Learning. Instead, it offers guidance on how to implement Impact Measurement at the program level and will in the future include guidance on how to manage information for the purposes of the IM system.




**Some challenges and thorny issues**

We recognize that the work COs do is not without challenges and struggle. Therefore, in this section we address some of those challenges and offer support by outlining several thorny issues that have come up in our conversations at regional program conferences, in our meetings with Program Teams, and elsewhere.

***How does impact measurement deal with breakthroughs in theories of change (TOCs)?***

One of the elements of a program TOC we have discussed in CARE is a breakthrough. A breakthrough[1] is a change that represents a significant leap forward that is not easily reversed. A breakthrough represents a change that affects both the breadth of impact (increasing impact on many more people in our impact group) and the depth of impact (increasing the level of wellbeing or transformation in the lives of our impact group). The change resulting from a breakthrough is reflected in the lives of people in our impact group whom we directly work with, as well as people in the impact group outside of our operational areas.

As breakthroughs offer the opportunity to significantly increase impact, it is important that we recognize these opportunities, whether they are achieved by CARE or through the efforts of others. An important element of the concept is that a breakthrough does not serve as a breakthrough unless we recognize it as one, and act on it. In doing so, it makes us more aware and attuned to recognizing other breakthroughs that occur, especially unanticipated ones, and enables us to join forces with others to realize their potential.

***Types of Breakthroughs***

*Structural or Policy Change*

Examples here include a change in land inheritance policy that opens the possibility for women to inherit land, or decentralization processes that make it easier for communities to influence state run agricultural programs.

*Change that Sets Precedent*

Something positive that occurs for the first time and sets precedent that opens the possibility to replicate multiple times at scale, e.g. the first time a women farmer's group challenges injustice in land allocation practice and wins.

*Critical Threshold of Incremental Change*

This could be incremental change that reaches a certain level of threshold from where it will be impossible to go back to the former state. It indicates a nearing of critical mass where practices or behaviors will soon be normalized. For example, in over 70% of communities in our operating area, local authorities involve women's groups in their budgeting process.

For all of these reasons, it is very important that we try to identify breakthroughs and work with others towards achieving them in the program design and implementation. However, given their very nature, breakthroughs cannot be a central part of the program Impact Measurement system. Most breakthroughs are one-time events (structural, policy or precedent change) that really do not require measurement per se, nor does measuring the breakthroughs tell us much about how they are being leveraged to bring about deeper and wider change the lives of the impact population. They are also, by their nature, often unpredictable as mentioned above.

Breakthroughs play a role in the IM system in two areas: the critical threshold of incremental change breakthroughs; and the measurement of leveraging the other breakthroughs for wider and deeper impact. In these areas we can set indicators in the pathway, domain and between-domain levels (more on this below) that allow us to measure changes over time and our contribution to that change. However, it is important to understand that measuring whether a given breakthrough occurred or not cannot be all there is to impact measurement. More on this below.

***How does impact measurement utilize CARE's MDI+ approach?***

At the Istanbul 2008 conference, which officially started the program shift at CARE, the CI-wide group that met there decided that CARE would adopt an "MDI Plus" approach with the program shift. Adopting an "MDI Plus" approach means using the Millennium Development Indicators (MDIs), which reflect the Millennium Development Goals (MDGs), to measure and communicate CARE's contributions to social change.

Using an MDI+ approach to measure the results of our programming enables us to communicate our impact based on a framework that is externally recognizable and internationally agreed upon. The MDIs are accepted and used by peer INGOs, governments, development research institutions, international

[1] Categories of breakthroughs taken from draft document on breakthroughs, March 2010, by Andrea Rodericks.

development agencies, donors, and many others. The MDIs are used in government national development plans, in donor framework agreements, and in much research. Using this common framework of indicators allows us to communicate succinctly and powerfully what our contributions to development have been. The MDI+ approach eliminates the need for us to provide lengthy explanations of what CARE-specific indicators and "CARE lingo" mean. In short, the MDI+ approach gives us a shared language with those we seek to make our partners.

At the same time, the MDIs are not without some shortcomings. They do not adequately capture some areas that are central to our work. For instance, their sections on women's empowerment and governance do not allow us to capture CARE's experience and commitments in these areas. That is why our approach is MDI Plus. What does "the plus" refer to?

First, "the plus" in the MDI+ approach refers to enriching the existing MDIs with a few indicators developed out of CARE's experience. Our programming on and learning about women's empowerment has produced some indicators that are now included in our menu of global outcome and impact indicators. Similarly, CARE is doing some conceptual work on its governance experience, seeking to develop governance indicators on outcome and impact level. These additional indicators will enrich our work with the MDIs and allow us to better measure and communicate our contributions.

Second, the MDIs are most often measured on the national level. We have national-level statistics on maternal mortality in Peru, or national-level statistics on proportion of the population below $1.25 a day in Liberia. However, we know that our impact groups, being the most vulnerable and marginalized, are often excluded from many benefits of national development. They are often excluded from the national-level statistics advertised by some governments as well. Under the program approach, we make a commitment to measuring our success against the change that happens in the lives of our impact groups: those most marginalized, excluded, and kept from benefits and rights. Thus, the second aspect of "the plus" we add to the MDIs is to apply them to the most marginalized groups. By showing gaps that often exist between national-level accomplishments and the situations of the most vulnerable groups, we shine a light on exclusion and inequity. The MDI+ approach includes applying our metrics to the most vulnerable so we can hold power holders accountable.

What about CARE's Unifying Framework of Human Conditions, Social Positions and Enabling Environment? Don't the MDIs only reflect human conditions changes and thus leave out our work on social positions and enabling environment change? The indicators that have been labeled "impact indicators" in the MDI framework do in fact focus on human conditions change. They do not explicitly reference any changes on the other two levels that we would aim to facilitate via our programming. However, the ultimate change we seek – the ultimate accomplishment of our long-term impact goals – is an improvement in the human conditions of our impact groups. A social position change or an enabling environment change is only meaningful if it leads to a better life for those whose lives we seek to help make better. This means that our ultimate impact should be measured by impact indicators on the human conditions level, and that our work around social positions and enabling environment should be measured on the outcome level.

For example, we can have an environment in which emergency obstetric care is provided for all women who seek it. And we can have a community which considers the lives of mothers precious enough to spend resources on emergency obstetric care. But unless these two outcome-level changes lead to the human conditions impact of having fewer mothers die in childbirth, they are not meaningful in themselves. Thus, we use outcome-level indicators to measure social positions and enabling environment changes, then demonstrate how these lead to impacts on the human conditions of our impact groups.

But what about the fact that our global outcome indicators do not directly reference social positions and enabling environment? For example, how does an indicator like "primary school completion rates, by sex" tell us anything about social positions and enabling environment? To discuss this question, we will introduce the concept of proxy indicators. Proxy indicators are indicators that help us register a sign of change in phenomena that cannot be measured directly. In our example, in order to see as many girls complete primary school as boys, we would have to first see a change in how much families, communities

and societies value girls' education, girls' ability to make a contribution once educated, and the importance of having better opportunities for girls through education. This change in social position is complex and difficult to measure, like most social changes. The indicator of "primary school completion rates, by sex" is a proxy indicator for this social position change: it allows us to register what change would need to have already occurred in order for girls to complete primary school at the same rate as boys.

The same is true for enabling environment: unless there is an environment, created by good governance, provision of quality services such as primary education, access to these services, etc., girls (or boys, for that matter) of the most marginalized communities will not have high school completion rates. Thus, if we measure change on the proxy indicator "primary school completion rates, by sex," that will signal to us that change has already occurred in the enabling environment as well. In short, we have outcome indicators which are proxies for complex changes in social positions and enabling environment. Now, what we need to focus on is doing the analysis that will reveal what a number measured against a proxy outcome indicator means and what social change has already taken place. More on analysis in Chapter 3.

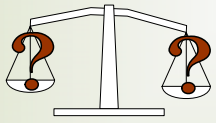
***What about indicators for impact measurement? Aren't indicators the core of impact assessment?***
To be sure, selecting appropriate indicators strategically is key to any level of measurement and assessment, and the same holds true for the impact level. In light of this, we offer some basic approaches and rules of thumb to indicator selection, below, then a more extended discussion in Chapter 3. However, we want to emphasize that IM is about analytical thinking more than about anything else: being reflective, critical, analytical thinkers about our work. We need to focus on what we need to know, what we need to learn, what questions we need answered in order to improve how we do our work and thus maximize its impact. The core of IM is not measurement, or indicator selection, or how to do large surveys or any other sorts of data collection. The core of IM is analytical, critical thinking. To be sure, there are technical aspects to IM, and using the appropriate metrics with a rigorously applied methodology is important. In FY11, the guidance in this document will be expanded to provide guidance, standards, technical discussions and so on, on the technical aspects of IM. However, the most important process of IM is to think analytically about what we are doing and what questions we need answered. Once we do the analytical thinking – and only then – indicators and methods will suggest themselves.

With this clarity in place, we now review the basic understanding of impact indicators that will guide our thinking throughout the IM work. Practical examples of these rules of thumb from CARE's current work will be provided in subsequent chapters.

Most of us who have worked on measurement can agree that working with indicators – finding the right ones, choosing between different options, developing new ones – is important but difficult, not to mention frustrating at times. In this section, we will discuss the challenges of indicator selection and set out some approaches or "dos and don'ts" to make our task easier. For now, we will focus on how to generally approach indicator work; in FY11, we will develop detailed technical guidance including operationalizations of specific indicators, tools for data collection and the like.

*(a) DO clarify conceptually what you need to measure first, DO NOT go directly to indicator selection before doing the thinking*
Often, when we talk about measurement, we immediately jump to indicators—but this is not a good approach. We can only select appropriate indicators and be strategic about our selection once we are extremely clear on what we need to measure. We need to do our conceptual analytical thinking first. Once we have identified a concept or an issue, we need to probe and ask questions about it until we have come to the bottom of what it stands for, what its underlying components are, what a change in it would tell us, and what changes we would like to see. Only then should we start the process of selecting an indicator to measure it.

For example, imagine that we are working with a vulnerable group of people who are exploited by money lenders who give them loans at very high rates.



**Think Conceptually!**
Know what you're measuring before you select an indicator with which to measure it.

We need to find a good indicator to capture this sort of exploitation and dependence.

Clarify the underlying problem first. Our first task – before considering any indicators at all – is to conceptually clarify the underlying problem that needs to be captured. Our conceptual thinking may go along these lines:

- Is the problem that these people borrow money? No, borrowing money in itself is not necessarily a negative thing. If, for example, a father borrows money at a good rate to pay for the treatment of illness in the family, and then is able to repay the money quickly, the borrowing of money in itself is not a bad thing for him.
- Is the problem that a vulnerable person has no access to a bank or similar financial institution to which others have access, but instead has to go to the village moneylender who offers exorbitant rates? No, because if the bank offered the loan at the same exploitative high rate as the local moneylender, the bank would be no better. It is the exploitative rate, not the source of the loan itself, that is the underlying problem.
- Is the problem the rate of loans, or that vulnerable people only have access to the very exploitative rates, whereas other (less marginalized) people can get a loan for a lower rate? Yes, when a group of people can only get a loan at an extremely high rate and is being denied access to loans at lower rates than less marginalized populations can get, this group is being exploited.

Once we have gone through this conceptual thinking, we can then select an indicator for exploitation: *the interest rate on loans*. Note that this very simple indicator already exists, and is widely used and well tested. Had we skipped the analytical steps above to get clarity on the underlying issue, we could have easily decided, for instance, that the problem is "exploitation" and started designing some very complex indexes to measure it. Instead, solid conceptual thinking gave us clarity on the underlying problem we need to measure, which then led us to a fairly simple existing indicator that we can use: *interest rate on loans.*

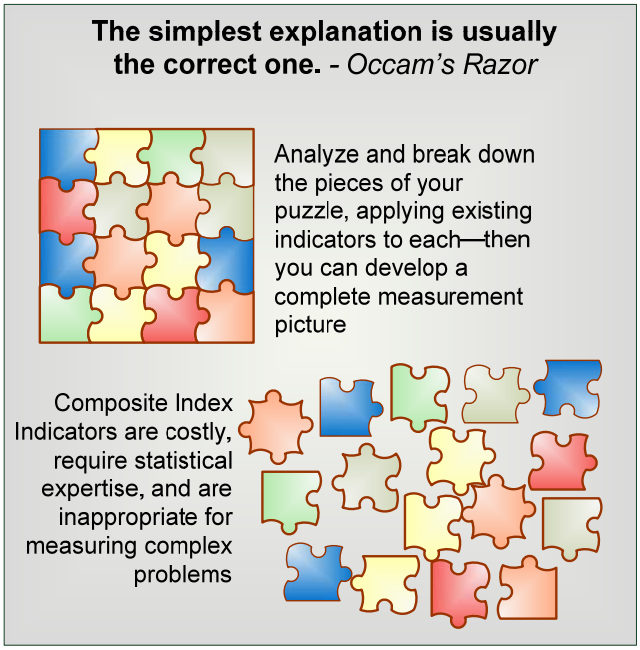
*(b) DO use existing indicators, DO NOT invent new ones*
Coming up with good indicators is a difficult task that requires various technical skills and expertise. It requires the person designing the indicator to be familiar with methods to operationalize concepts; methods to weigh variables in parsimonious models; and so on. Because CARE is not a research institute but an implementing development organization, we often do not have the expertise—nor do we need—to design indicators in-house. Instead, we should employ existing, proven indicators developed by experts, rather than try to invent new ones ourselves. If we use existing indicators, we will benefit from the expertise and experience of those who know how to develop them and have already done the work; we will save ourselves the trouble of having to test the indicators; and we will avoid having to provide lengthy explanations and justifications when communicating our work externally.

Sometimes, we are so focused on the details of our own work that we think we cannot find existing indicators to help us measure what we need. However, that is rarely true. If we make a commitment to exhausting the options of existing indicators first, we will not need to invent new ones—and, ultimately, our work will be much higher quality and defendable as a result. For example, assume that we are working with an impact group that is exploited by selling advance labor during seasons when the price of labor is low, then having to perform the labor during seasons when its price is high. This exploited group obviously does not benefit from the seasonal increase in the price of labor, and we are interested in measuring this economic exploitation. We could start developing a very complex index, or a composite indicator, including various aspects of economic exploitation. Or, we could select a very simple indicator that has been already developed and tested numerous times by economists: *seasonal wages*. We need to measure wages in the different seasons for our impact group and for a group that does not experience the same exploitation (e.g., nationally, for a higher caste, etc.). The difference in wages between those who sell their labor expensively when its price is high, and those who sell it cheaply even when its price is high, is an excellent indication of exploitation.

*(c) DO use simple indicators, DO NOT think that a complex concept requires a complex indicator*
Like many answers to questions, often the simplest makes the most sense. Similarly, the best indicators are the simplest. They tend to be straightforward, easy for everyone to understand, and minimize the

potential measurement error. They also enable us to focus on analysis and learning rather than on complicated measurement. If we encounter a complex concept or problem, as we often do when we work on marginalization, empowerment, exploitation, rights and so on, we need to break those complex concepts down until we can find simple indicators for them. It is always better to simplify a thought and attach a simple indicator to it than to try to design complex indicators to measure complicated thoughts.

Composite indicators are inappropriate for analyzing complex processes. When we are dealing with a complex problem, such as economic exploitation or marginalization, we are tempted to design an index to measure this problem. In general, indexes are useful for predicting macro-level trends. For instance, if the government wants to predict the movements of the economy in the next few months, it will have its statistical institute calculate a composite index of many different indicators, including employment, manufacturers' production, value chains, financial changes, interest rates, and consumers' behavior. This approach is not appropriate for our purposes.




To continue with the previous example about vulnerable people exploited by village moneylenders, imagine that we wanted to measure economic exploitation. This is a very complex and multi-faceted problem. Instead of doing the conceptual thinking to simplify what we need to measure, as we did in the example above, we could start to develop a very complex indicator for economic exploitation, which would likely take the shape of a composite index. A composite index of indicators combines various indicator components to produce one overall number that can be tracked over time. Our index of economic exploitation could contain an indicator measuring whether those borrowing money are extremely poor; one that measures whether they s ell advanced labor; one that measures whether they are aware of the prices of labor each season; and still another that measures whether they take loans from the village moneylender. We could then attempt a complex calculation to aggregate all these variables into a composite index measure. As the following explains, using composite indexes of indicators is not a good approach for our work for a variety of reasons:

- First, a composite index approach requires statistical technical expertise that we do not have to conduct the work rigorously.
- Second, the approach requires a lengthy and expensive measurement.
- Third, the approach would not tell us anything that a simple indicator like the one discussed above – interest rate on loans – does not already tell us.
- Fourth, composite indexes are not useful for explaining complex processes, such as what we do when testing our TOCs and analyzing our contributions to social change. If one thinks of indicators as once-removed from reality (e.g., there is "real life," then there is an indicator to reflect real life on paper), then indexes are twice removed from reality--think of them as reflections of reflections.
- Fifth, constructing a complex indicator such as an index is difficult, demanding statistical work – one cannot simply decide to add a few indicators together. Even the selection of the components, which will be part of the index, needs to be tested statistically. The calculation of measures to produce the final number on the index is difficult statistical work as well.

In summary, complex indicators such as composite indexes are to be avoided. If you must use one, do so with extreme consideration and care, and always use an existing one. However, it is best to analyze and break down a complex problem or issue to its component parts, then assign a simple indicator to each part. A complex problem is best captured and analyzed with simple indicators.

*(d) DO use one indicator per change you seek to measure, DO NOT use multiple indicators for the same concept*

If we survey the different areas in which we work, we will find hundreds of indicators for each. For example, we can find numerous indicators for education: enrollment rates, completion rates, ratios of teachers to students in the classroom, and many others. This does not mean, however, that we need to measure all of them. Using as many indicators as we can think of in our data collection tool does not improve our measurement; it only makes it longer, more costly, and more burdensome. Instead, we should focus on selecting one good indicator for the underlying change we want to see, and concentrate our resources on measuring and analyzing that.

For example, our global menu of outcome and impact indicators lists five different indicators in the area of Environmental Sustainability. Our job is to figure out which one would best capture the change we are working towards in our program. We do not need to include more than one of those five indicators. A good detailed analysis of one of them will go further than lengthy measurement on all five.

*(e) DO consider financial, time, and skill implications of indicators, DO NOT be afraid to choose the cheaper, easier to measure ones*

When we have gone through the four *do*s above (i.e., done our conceptual thinking, identified potential existing indicators, thrown out the complicated and selected the simple ones, and tried to narrow them down to one ), we are sometimes still left with choices among good indicators. How do we select one among many equally good options? We should seriously consider indicators based on whether they are less expensive to measure than others, would take less time, or would require fewer specialized technical skills.

**Less can be More!**
When considering equally applicable indicators, go with the cheapest, fastest, and easiest indicator.


For example, imagine that we need an indicator for food security and that we have two options: *caloric intake* and *dietary diversity*, both disaggregated *by sex*. Which one do we choose? Caloric intake is extremely difficult to measure and calculate, as it requires difficult precise recall on the part of survey respondents. It also requires an expensive survey done by people with special technical skills to gather the data and do the necessary calculations. And it would also require a long lead time to organize the resources and people needed. Dietary diversity, on the other hand, is a good proxy for food security: people who have higher dietary diversity are less threatened by food insecurity, because those who diversify their diet over time have escaped the threat of food insecurity. Dietary diversity is also easier to measure for both survey enumerators and respondents: the recall involved here is much less demanding than would be the case in trying to calculate calories. Because this is a simpler indicator, it would also require less skill, time and money to measure. Based on these practical considerations, we choose dietary diversity as our indicator, and leave the measurement and calculation of caloric intake to the DHS and other national surveys done by research institutes.

*(f) DO focus on analysis, DO NOT think that an indicator, however good, is a substitute for good judgment and good thinking*

Indicators are designed only to facilitate our analysis and thinking, not to replace them. Even if we select the perfect indicator but then do not analyze the data we collect on it, reflecting on what it is telling us, then we have not done our job. Indicators are not short cuts for thinking. They are simply a tool to help us collect the data and do the analysis that should inform our thinking.

Further, indicators only indicate. That is, they point in the direction in which change has taken place. They are not a substitute for our work – we need to do the analysis of whether the change is positive or negative, expected or unexpected, as described by our TOC or not, large or small. Indicators are good for

data collection, but the core of our task with IM is analysis. We have not done IM until we complete our analysis, interpret the information derived from the analysis, and draw out its implications for our work.

To illustrate how *collecting* data on an indicator is different from *analyzing* and *interpreting* data, consider the following example:

Imagine a project to reduce violence against women and girls, and that some of the main interventions include raising community awareness and encouraging people to report instances of such violence and seek legal redress. The *number of instances of reported violence against women and girls* could be an output indicator for this project. It is very likely that, between the baseline and the mid-term, the number of instances of reported violence would increase; our data collection would show the numbers going up. If we simply collected the data and ended our effort before properly analyzing those data, we would have no idea what this increase meant. It could mean that *more* violence against women and girls is taking place, or it could mean that the amount of violence has not changed, but the *increased community awareness* means more instances of violence are actually reported, rather than going unrecognized, as they did before awareness was raised. It could also mean that the amounts of violence and awareness have not changed, but more people had been empowered to *report* instances of violence. We cannot know which of these changes has taken place simply by saying that the number of instances of reported violence has increased. We will understand the change that has taken place only if we analyze our findings in their context, compare them to findings on other indicators, and draw conclusions from this analysis.

Having a good indicator does not, in and of itself, replace taking the time to analyze the collected data, which reveal very little without an analytical context. The key, therefore, is to do the in-depth analysis to understand the meaning of the findings and act accordingly.

Having introduced impact measurement systems and some basic rules of thumb that will guide us in building them, we now turn to the key processes of IM systems and to some more technical data issues. Chapter 2 discusses the main processes of an IM system: testing theories of change (TOCs) and tracking progress toward impact goals. Chapter 3 deals with data collection, management, and analysis. While all these chapters present a more conceptual discussion of the issues, they also contain concrete, practical examples drawn from the experience of various CARE COs. In FY11, these examples and guidelines will be expanded with operational information.

# II. ANALYSIS AND INTERPRETATION: TESTING THEORIES OF CHANGE AND TRACKING PROGRESS TOWARD OUR IMPACT GOALS

[insert text from Maliha, conceptual explanation of what the main processes in an IM system are. First conceptual treatment, then step-by-step guidance – PENDING]

**Step-by-step guidance: how to carry out the key processes of impact measurement**

Having already discussed these processes in theory, we now turn to some guidance on how to carry them out. The remainder of this chapter is dedicated to specific, step-by-step instructions and illustrations of the key processes of testing the TOC and tracking changes toward the impact goal.

Before we discuss testing the TOC and tracking changes toward the impact goal, let us define a few terms we can use.

A *theory* of change describes a process of desired social change by making explicit the way we *think* about a current situation or problem, its underlying causes, the long-term change we seek, and what needs to happen in society in order for that change to come about. A theory of change helps us uncover the thinking that guides our interventions/actions. By making explicit our thinking, we can be proactive in testing and adapting this thinking, which, in turn, helps us improve our interventions/actions over time (more on this on the p-shift wiki, http://p-shift.care2share.wikispaces.net/Theory+of+Change+Guidance). In sum, a *theory* is a conception, or our best guess, of how certain social changes can come about. It is a 'best guess' in the sense that while it is generally based on programmatic experience and knowledge, it is yet to be tested and verified.

> A *theory* is our thinking on how social change will happen. It consists of numerous *assumptions*.

When we design a TOC, we make a number of *assumptions* about the social change processes that we envision happening. An assumption is a supposition or a specific guess we make about a given social change process. Our TOCs are made up of assumptions. The entire set of all our assumptions about social change and impact makes up our TOC. For example, one assumption we may make is that better economic opportunities for women result in less domestic violence against women. When designing our TOC and a domain of change around the economic empowerment of women, we may assume that economic opportunities for women will help reduce the violence in the home against them. Sounds plausible. However, we may not yet have the solid proof that this is the case, or that the process happens just the way we think it would. In order to see whether our assumption was valid, we would have to test it.

> An *assumption* is a supposition or guess about a given social change process. Our TOCs are made up of *assumptions*.

A *hypothesis* is an assumption which we actually test. That is, we will make numerous assumptions when designing a TOC. Some of these will be key to our theory and will need to be tested; others will be less important and will remain assumptions. When we take an assumption and state it in a form that allows us to test it – and then either prove it, or disprove it – we have turned that assumption into a hypothesis. We can make a very general assumption about something. However, if we want to test it, we have to make it very specific and state it in a certain way in order for it to be testable and disprovable. Many specific hypotheses can be generated out of one general assumption, as we will see in the example below. Most commonly, hypotheses are stated as *if-then* statements. This simply means that we have specified what needs to happen (in the first part of the sentence) for something else to happen (in the second part of the sentence): *if* this happens, *then* that will happen. To use the example above again:

> A *hypothesis* is an assumption which we actually test. A *hypothesis* is an assumption that is able to be tested and either proved, or disproved.

We may take the

*Assumption: Better economic opportunities for women result in less domestic violence against women.*

And turn it into these

*Hypotheses: (1) If there are more economic opportunities for women, then more women will participate in them; (2) If more women participate in economic opportunities, then their contribution to household income will increase; (3) If women's contribution to household income increases, then they will be more valued by members of their households; (4) If women are more valued by members of their households, then the domestic violence against them will decrease*.

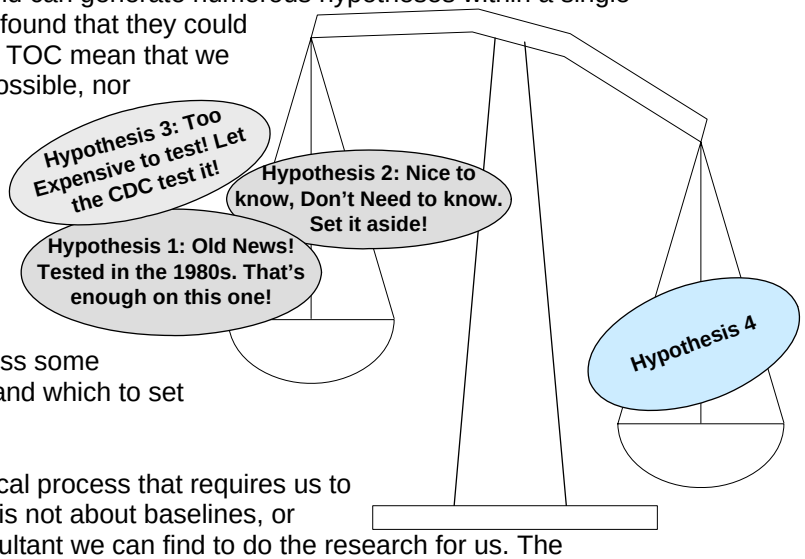
We may come up with many more examples, but the point is that we will make many guesses and suppositions when designing our TOCs (these are assumptions), and we will test some of them (these are hypotheses) in order to assess whether our thinking was valid.

Armed with these terms – theory, assumption and hypothesis – we will now discuss a step-by-step process for testing a TOC and for tracking changes toward an impact goal.

***PROCESS 1 – Testing the TOC – Some Ideas and Approaches***

As we said, a theory of change (TOC) is a set of assumptions and hypotheses about how we believe a desired social change will come about. Our TOC is based on experience working with communities, on knowledge we have accumulated over time, on what others have learned and shared, etc. In this way, it is not random and the suppositions and guesses – or assumptions – we make in it are not baseless. However, we still need to test them in order to ascertain their validity and ensure that our thinking about the specific processes of social change we have described is accurate. Testing the TOC, then, is the process of making our assumptions explicit; turning them into testable propositions (called hypotheses); doing some research to test these hypotheses; and analyzing what the findings imply for our TOC overall.

To be sure, we will make numerous assumptions and can generate numerous hypotheses within a single TOC. When some COs examined their TOCs, they found that they could generate hundreds of hypotheses! Does testing the TOC mean that we have to test them all? No, it does not. It is neither possible, nor desirable for CARE to test hundreds of hypotheses. We will learn how to strategically select a few out of the many hypotheses to test. Not all hypotheses are born equal: some do not need to be tested because they have already been tested in the past; others are not so key and vital to our work, so we can leave them aside for now; yet others are just too expensive for us to work on testing with methodological rigour. So, we will discuss some criteria for how to decide which hypotheses to test and which to set aside.




Finally, the process of testing the TOC is an analytical process that requires us to do critical thinking. The process of testing the TOC is not about baselines, or how much data need to be collected, or which consultant we can find to do the research for us. The process of testing the TOC is mainly about critically examining our own thinking, being clear about our own assumptions, and, once the research has been completed, figuring out how we need to adjust our TOC and our actions based on the findings. Ultimately, then, we test our TOCs because we need the findings from the testing to improve our program quality, ensure we reach the impact goal, and remain accountable to those implicated in our TOC: the impact group.

To test the TOC, we follow these steps. A detailed example of them is provided shortly.

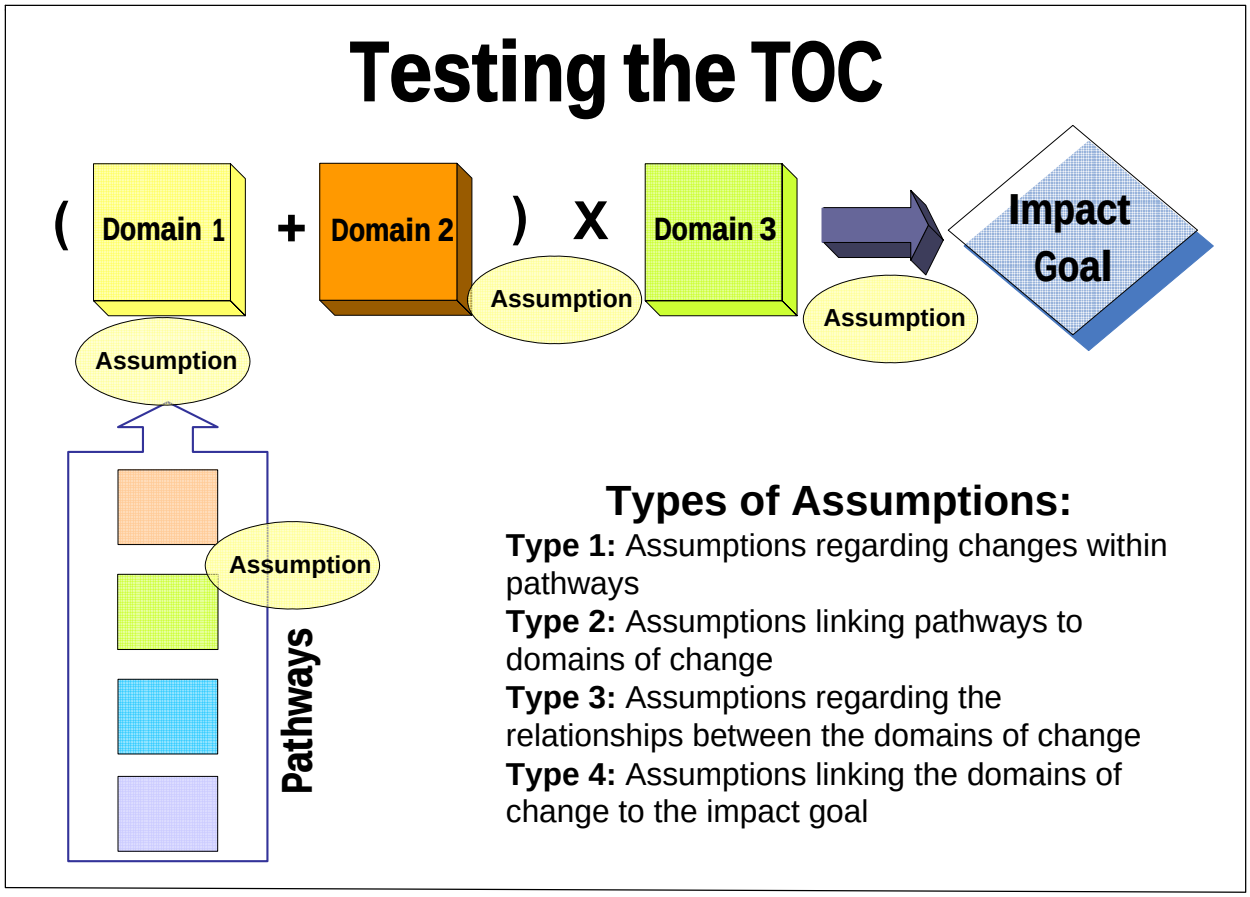
► *Step 1:*
*List all hypotheses that are part of the TOC, and identify their type:*
→ hypotheses within pathways
→ hypotheses linking pathways to domains of change
→ hypotheses between domains of change
→ hypotheses linking the domains of change to the impact goal

The main goal of Step 1 is to make explicit the thinking we have embedded in the TOC: to spell out what we might have assumed about various social changes and how they would happen. For instance, if we have assumed that, generally, more economic opportunities for women have something to do with reduced levels of domestic violence, now is the time to make all those thoughts explicit and to detail them. Then, we state them as testable, disprovable *if-then* statements – that is, we turn them into hypotheses.

Further, there are various levels of the TOC at which we make assumptions and for which we can come up with hypotheses to test.

- One type of hypothesis is about the changes we describe *within the pathways*, specifying steps needed for a sequence of social changes to occur.
- A second type of hypothesis is the one *linking pathways to domains of change*. Here, we explain how the various changes within a pathway contribute to changes in the domain.
- A third type is a hypothesis that specifies the *relationships between domains of change*. For instance, many CARE programs have TOCs in which governance is a domain of change and it has a multiplier effect on all other domains of change. The notion that good governance can multiply the effects of changes in other domains is an assumption – it needs to be turned into a hypothesis and tested.
- A fourth type of hypothesis tells us how all the domains of change are linked and work together *toward the impact goal*.

The diagram below represents a typical TOC and illustrates the different types of assumptions and hypotheses we can have:




Again, the main goal of Step 1 of testing our TOC is to explicitly spell out the assumptions we have made when designing our TOC. In doing so, we need to pay attention to all the different levels where assumptions are made: within pathways, between pathways and domains, between domains, and from domains to impact goal. The goal is to list all these assumptions, and challenge ourselves to be explicit and clear about what we may take for granted as true. Once we have listed our assumptions, we can state them as hypotheses for testing.

Then, we can proceed to Step 2:

► *Step 2:*
*Prioritize hypotheses to be tested, based on the following criteria:*

→ has the hypothesis been tested before, by CARE or by someone else?
→ is the hypothesis strategically important to CARE's work?
→ is testing the hypothesis feasible given our resources (financial, skills, time, etc.)?
→ document your process of selection of hypotheses to be tested – this becomes part of your arguments and communications to donors, partners, CARE, and others

As we discussed before, we make very many assumptions and can therefore generate just as many, or more, hypotheses. Do we need to test them all in order to say that we have tested our TOC? No. We do not. What we have to do is strategically select a few of the many possible hypotheses that we need to test.

How do we make that selection? How do we know which hypotheses to test? We can make the selection by answering a few simple questions and finding out a bit of information about what work others have already done before us. To prioritize hypotheses for testing, we answer these questions, in this sequence (designed to help us eliminate hypotheses we do not need to test):

- *Has this hypothesis been tested before, by CARE or by others? If yes, set the hypothesis aside and simply refer to the sources of information you found. If no, go to the next question.*

For example, if we have a program on maternal and child health, and one of our assumptions has to do with the benefits of breastfeeding for child health, we do not need to test a hypothesis based on this assumption. There is a large body of literature that tells us about the health and other benefits of breastfeeding – we do not need to replicate it.

In the work we do, there is a good chance that a given hypothesis has already been tested, either by CARE in previous work, or by others. There is a large body of knowledge on the work we do produced by other development NGOs, universities, research institutes, government-funded agencies, etc. Also remember that CARE is a source of information as well: other projects, other programs, other COs may well have tested what you are working on. We should never assume that something has not been tested. Instead, we should do our homework and find out some information: has this been tested? What does existing knowledge offer us on this hypothesis? It is always better to invest some time in investigating what has been tested before and what is available to us than to invest resources in re-inventing the wheel.

- *If the hypothesis has NOT been tested, is it vital or strategically important to CARE's work? If no, set it aside and simply state that it is a hypothesis to be tested, but not of strategic importance to our work. If yes, go to the next question.*

Some hypotheses are simply more important to our work than others. We may hypothesize many things, but they will not all be equally important to our work. We are only interested in testing those that are truly crucial to what we do, in this program or as an organization, and we cannot leave it to chance that someone else will test them. By testing those critically important hypotheses, we are adding to the shared knowledge on a given issue something that no other actor will add. If this is not true, then we should not test the hypothesis.

- *If the hypothesis IS of strategic importance to our work, is it plausible that we will be able to test it? That is, do we have the resources to test it in a way that will produce solid, credible evidence: money, time, technical skills, human resource base to draw from (staff, consultants, research partners, etc.)? If no, set the hypothesis aside and explain why it is so important but what resources are lacking for us to test it well. If yes, go to the next question.*

This step is designed to keep us realistic about what it will take to test each hypothesis that we think is strategically important to our work. Every time we test a hypothesis, we need to make sure we have the resources to test it well – that is, to use the right methods in the right ways so that the evidence we produce is solid and believable by others. Testing for the sake of testing is not a good use of our resources and should not be done.

Be very realistic about the resources that testing a given hypothesis requires. Some hypotheses will be more expensive to test than others. For example, some will require large quantitative surveys to test, which are expensive, time-consuming, and require very specific technical skills. Other hypotheses can be tested using methods and approaches with smaller resource implications. More on this below. For the moment, the point is, this does not mean you should never test expensive-to-test hypotheses – just that you should be very realistic and practical, and make a strategic decision.

- *Document, document, document!*

By going through these questions in Step 2, you have just engaged in a very challenging and sometimes labor-intensive but valuable analytical process. You have made your implicit thinking explicit and clear; you have spelled out what about your work needs to be tested and verified; you have investigated what existing wisdom has to offer on testing and verifying your hypotheses; and you have made some tough strategic choices about what you can test and what you cannot test. In short, you have won half the battle! You have already done a lot of the conceptual thinking, reflection and analysis around testing your TOC.

Now document all these steps. Make sure you have the list of assumptions and hypotheses you did in Step 1. Record the type of each hypothesis so you know, for example, whether it belongs within a pathway or whether it links two domains of change together. Make sure you note down all the sources you find that have already tested some of the hypotheses. Makes sure you write down the analysis of what's strategic for us to test and of what resources we have or do not have to test it. Document all your thinking, just like you do when you design the TOC.

If you can document this work, you will have very powerful information to share with others. You can use it when making arguments to a donor about funding. You can share it with your partners or with those you seek to partner with. You will use it and refer to it over the 15-year lifespan or the program. You can send it to the rest of CARE when you present your work or seek assistance. Documenting your analyses and reflections when testing the TOC is vital.

To summarize, we will come up with many specific hypotheses out of the more general assumptions we have made when designing our TOC. However, we will not be able to, and do not need to, test all of them. We will answer a few questions about each in order to decide whether we should test it or leave it aside. The appendix to this chapter offers a decision-making flowchart summarizing the questions and decisions we would make when prioritizing hypotheses for testing. Review that flowchart before proceeding to Step 3 of testing the TOC:

► *Step 3:*
*Select an appropriate methodology for testing the hypothesis, based on the following criteria:*

→ what do you need to know? What specific knowledge is the testing supposed to produce?
→ what purpose will the knowledge be used for (for example, is this internal validation for CARE? Are you aiming to convince someone else? What does that audience find convincing?)
→ how do we get this information/what needs to be measured?
→ what information/knowledge already exists and how can we use it to answer our questions?
→ what are the possible relevant methods for coming up with the information/knowledge that does not yet exist?
→ what are the resource implications of each methodological choice above (human resources and skills, time, technical, financial)?
→ how do the methodological choices compare?
→ document your process of method selection – this becomes part of your arguments and communications to donors, partners, CARE, and others

Once we know which hypothesis we will be testing, we need to make some decisions about what methodologies we can use for the testing. The main goal of Step 3 is to make us think strategically about two questions:

"What are appropriate methodologies for producing solid evidence on our question?" That is, we should not immediately use a method simply because we have always used it before or because that is the only method we know. We need to consider what methods are available to us and which one will be best in this situation. Read more on this in Chapter 3.

"Of the possible appropriate methods, which one do we have the resources to carry out well?" That is, more than one method may be appropriate for answering our questions – but appropriate methods will differ in the resource demands they make on us. All else equal, we should select the method that we can do well, with rigour, adhering to the rules of that methodology. Read more on this in Chapter 3.

Here, Step 3 calls on us to:

- Be very clear about why we are testing a given hypothesis: We may need to verify, for ourselves, that our initial thinking around a given social change process was correct. So we ourselves – CARE – may be the audience here. Or we may need to convince others that our thinking and model of a social change process works. So we may have external audiences for whom we would like to produce convincing evidence. Having clarity on the audiences and what each finds convincing will help us decide what the appropriate methods of testing a given hypothesis are. One audience may only be convinced by large-scale quantitative surveys. Another audience may need to see more qualitative, case-study based evidence. We need to have this clarity before we embark on testing anything. Otherwise, we risk producing something which is not convincing and not useful with the audiences we try to use it with.
- Be very clear about what questions we need answered, or what we are hoping to learn from the testing. We need to formulate clear, concrete, specific questions that we need answered by testing the hypothesis. The purpose of testing a hypothesis is to gather knowledge, to answer questions that we need answered in order to advance our work. What are these questions?
- Be very clear about what knowledge already exists on our questions. We need to make sure we are not duplicating efforts, or reinventing the wheel, by doing research that others have already done. Our resources are limited and precious, so when we use them to do research, such as testing a hypothesis in a TOC, we should make sure this produces new knowledge and adds to what we know, not redoing work already done by others.
- Be very clear on which methods we choose, in an informed and strategic way. We have numerous methodological choices. Whichever one we choose to proceed with, we need to follow it very rigorously and carefully – otherwise, our evidence will not be convincing. Read more on method selection in Chapter 3.

Once we have achieved clarity on these questions and decisions, we need to document our thinking and process. Why did we decide what we decided? How did we approach it and think about it. This information becomes part of testing our TOC, part of our program documentation, and a very powerful part of what we communicate to others – inside and outside CARE – about our programmatic work.

Having documented which hypotheses we have prioritized and how we will go about testing them, we can proceed to Step 4, doing the testing:

► *Step 4:*
*Test the hypothesis, following the selected appropriate method rigorously*

You will immediately note that the step you may have assumed to be the most detailed – doing the actual testing – is actually the shortest in this guidance. We do not go into detail of how to apply different methods rigorously; we do not even talk about specific methods. Why not?

This is because the main goal here is to discuss the strategic choices and decisions that the program team has to make together, not to delve into how to implement research techniques. Not all of us are research scientists, and not all of us will be involved in the details of the research of testing a hypothesis. In fact, CARE is not likely to have the research expertise to carry out the testing in-house. In most cases, we will need to partner with others – universities, research centers, other development agencies, etc. – or hire a consultant with specific technical skills to do the actual testing research. There will be research partnerships we need to enter into, or specialized help we will need to hire. However, it is our program teams that need to make the strategic decisions about testing our TOCs: what is strategically important to test, how we will use the evidence from the testing, what resources we have to test, how we will set up a research partnership or other relationship to help us carry out the testing.

Read more about research partnerships in Chapter 3 before proceeding to Step 5 of testing the TOC:

► *Step 5:*
*Write analysis of the TOC, considering the following questions:*
→ what is the role of the hypothesis you have tested within your TOC? Why is it so vital to test and "prove"?
→ what does your test of this hypothesis say about your TOC?
→ what does it say about models you want to show are working and their scaling up?

In Step 5, our entire process of testing the TOC bears fruit: we get to understand what the testing has shown us and what that means for our TOC and our work. Having tested an individual hypothesis or two, we now need to bring the results and findings back to the TOC and understand:

- What does the result of the testing tell us? Were we correct in how we thought the social change process would work? If yes, how do we proceed to incorporate and use this knowledge? If not, how do we need to revise or refine our TOC to include the new information?
- If we were testing to demonstrate that a model works and can be scaled up, does the model work or does it not? Can it be scaled up? How do we incorporate this information into our TOC

This step of testing the TOC invites us to do analysis: to think reflectively and critically about what we have learned by testing a hypothesis, what this implies for the TOC and for how we go forward with the program, and how we can communicate this knowledge to others. This thinking is again the job of the program team, not of the consultant or the person coordinating the logistics of doing research. This is a moment we can take as a team to reflect on the work we are doing in this program. And again, it is crucial that we document our thinking and our conclusions reached on the basis of testing the hypotheses we had prioritized.

Since we mention learning based on the findings of testing our hypotheses, let us make sure that we understand the relationship between testing the TOC and our learning agenda. The learning agenda is much broader and requires many more questions answered and sources of information tapped than testing the TOC can offer. We test the TOC to learn what works, what does not, and how we need to adjust or refine our program theories and strategies over time. The knowledge we derive from testing the TOC needs to be incorporated into our existing and future programming. The learning agenda, on the other hand, calls for us to accumulate learning beyond an individual theory that we can test. For example, our learning agenda will contain learning questions on partnerships, on contributions to broad movements for social change, on positioning CARE and building relationships based on shared commitments to certain impact groups or social changes, and much more, in addition to questions around TOCs. In this way, testing TOCs feeds into our learning agenda and supports one very important aspect of it, but it cannot replace our overall learning agenda.

A final point about testing the TOC, before we go through a detailed example to illustrate: in what sequence should we test the hypotheses? Should we start with the "big picture one" about the impact goal, then make our way down the levels of the TOC? Or should we first test all the ones within pathways, then proceed to the level of domains and so on?

There is no one answer to this question or any particular order that we have to follow to test the hypotheses. The answer will vary depending on the situation in each CO and with each TOC. For example, if you have a particular hypothesis that is really crucial for how you proceed with your program (for example, many of your strategies and approaches are organized around it), but there is so little evidence that you cannot proceed, then you may want to consider testing that hypothesis first. You would not want to go ahead with a TOC based around a single hypothesis that remains unverified, nor would you want to invest resources into strategies and approaches based on something you don't know the validity of.

Furthermore, the CO can be opportunistic in terms of testing the TOC. If there is a planned evaluation for a project, which you are going to conduct anyway and there are resources available for that, you can look into ways to incorporate elements of testing the TOC into that evaluation. For instance, an annual evaluation may be planned for a women's empowerment project which has "engaging men and boys" as a strategy to empower women. At the same time, you may have designed a TOC for a program on women's empowerment and chosen men and boys' engagement as one of the major pathways. You should take the opportunity of that project evaluation to test a hypothesis from this pathway. This situation of testing a hypothesis may not reach all the way to testing how the domains of change contribute to the impact goal in your TOC. But it can certainly give you nice ways to test hypotheses within your pathways. Incidentally, this is what we mean when we ask, 'has this hypothesis been tested before, by CARE or by others?'

Having discussed the steps of testing the TOC conceptually, let us now consider a specific example to illustrate and further clarify the steps.

***PROCESS 1 – Testing the TOC – An Example from CARE-Burundi***

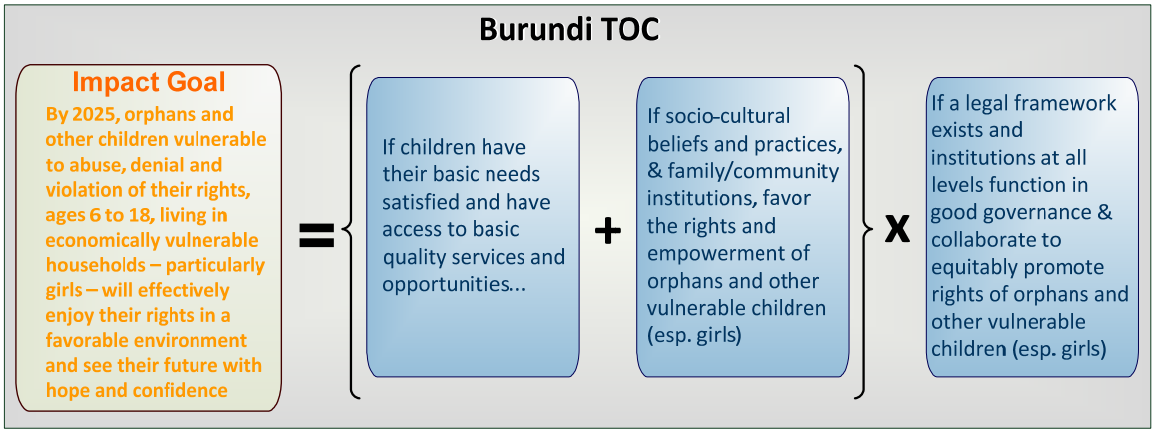
CARE-Burundi has identified orphans and vulnerable children as one of their impact groups. Their UCP analysis revealed three major underlying causes for orphan children's poverty and vulnerability:

(1) Limited access to basic quality services

(2) Socio-cultural beliefs and practices that discriminate against orphan children and prevent them from realizing their basic rights, and

(3) Lack of a legal framework and good governance at various levels (government, private, civil society) to promote and protect children's rights.

Based on these findings, CARE-Burundi developed the following TOC with three domains of change:



Below are the pathways that CARE-Burundi included in the TOC for each domain:

| | | |
|---|---|---|
| Economic reinforcement of households: food security, maximize agricultural production, develop non-agricultural revenue<br><br>Primary and secondary quality education (formal and non-formal)<br><br>Quality health system/ services improvement | Reinforce family and community capacities on their obligations to uphold and protect the rights of children<br><br>Promote the participation of orphans and other vulnerable children to take part in decisions that affect their lives<br><br>Transform socio-cultural barriers/promote positive values, beliefs and practices toward children | Promote good governance among institutions and CSOs at all levels for the equitable promotion of rights for OVC<br><br>Facilitate the collaboration of all levels of government institutions and civil society organizations<br><br>Revive traditional mechanisms for conflict resolution<br><br>Advocate for the adoption and implementation of OVC related laws |

Now let us see how the step-by-step process we outlined for testing the TOC would apply to this example.

***Step 1: List all hypotheses that are part of the TOC, and identify their type:***
In this first step, we take a close look at our TOC in order to start making explicit what we may have assumed when designing the TOC. This is a process of breaking down or elaborating our TOC. Our goal is to make assumptions explicit, then list them as hypotheses that we can test.

*Hypothesis 1:*
If the economic conditions of households are improved, then the percent of children who have their basic needs met will increase.
*Hypothesis 2:*
If the sources of household income are diversified by adding non-agricultural ones to agricultural production, then the overall economic conditions of households will improve.

*Hypothesis 3:*
If good governance at all levels promotes the rights of OVC, then this will multiply the positive effects of improving children's access to quality basic services and removing socio-cultural barriers to their empowerment combined.

*Hypothesis 4:*
If we combine improved access to quality basic services for children with lowered socio-cultural barriers to their empowerment, and see multiplying effects of good governance promoting the rights of OVC, then OVC will see their future with hope and confidence.

*Hypothesis 5:*
If OVC participate in decisions that affect their lives, then that will help lift the socio-cultural barriers to their empowerment.

These are some examples of hypotheses we may generate when unpacking the thinking behind this TOC. We need to be mindful that hypotheses are generated on all the levels of the TOC:

- Hypothesis 1 above links a pathway to a domain of change. It tells us how a specific pathway (the economic reinforcement of households) contributes to a specific domain of change (the basic needs of children).
- Hypothesis 2 above is within a pathway. It makes explicit an assumption, within the economic reinforcement pathway, that we have made regarding ways to increase household income.
- Hypothesis 3 above is of the type that explains relationships between domains. It tells us that the governance domain will interact with the other two domains by multiplying their combined effects.
- Hypothesis 4 above is of the type that links the domains of change to the overall impact goal. It tells us how the three domains will work together (two combined, the other multiplied) to contribute to achieving the impact goal.
- Hypothesis 5 above is again of the type that links a pathway to a domain of change. It tells us how the pathway of ensuring OVC's participation in decisions that affect them will contribute to the domain of changing the socio-cultural beliefs and practices around their rights.

Obviously, we have not listed all possible hypotheses to this TOC. There are many more that should be unpacked. However, these examples make the following points:

We need to be mindful of all the levels where we have made assumptions when designing the TOC and where we need to generate hypotheses. That is why we have four types of hypotheses in Step 1.

It does not matter which order we list the hypotheses and identify their type in. We can start from the big picture, impact goal level and work our way down to the details within each pathway. Or, we can start with the details of each pathway and gradually work our way to higher levels. The order does not matter. The important point is to unpack our thinking and make explicit all our assumptions so we can turn them into testable hypotheses.

Finally, we may notice that unpacking and making explicit our thinking on the TOC shows us that some of our previous assumptions may not sound so good anymore and we may now notice some gaps where the TOC needs to be further elaborated. When we design a TOC, we have the long-term, bigger picture in mind. When we start examining the assumptions and writing the hypotheses, we have a chance to look back, in more depth and in more detail, at our thinking about the social changes we desire. This is a normal process of refinement of the TOC. Thinking about the TOC in terms of testable hypotheses allows us to sharpen it. This is an iterative process: it sends us back and forth between the long-term, big-picture considerations when designing the TOC and the detailed, small-step elaborations when thinking about how to test it. In this way, the IM system helps us test but also refine our TOC.

***Step 2: Prioritize hypotheses to be tested, based on these criteria:***
Now that we have listed all our hypotheses, how do we know which ones to actually test? We need to prioritize some for testing and set others aside. To do this, we will ask some questions about each hypothesis:[2]

*Hypothesis 1:*

1. Has this hypothesis been tested, by us or someone else, before? No, we have found some research on similar questions, such as children's nutrition when the households they live in are economically strengthened, but not on the link between the economic situation of households and children's other basic needs. So, we proceed to ask other questions about this hypothesis.
2. Is this hypothesis crucial to our work? No; it is certainly important and figures in our thinking, but if we look at the OVC program, the crucial element we can contribute is around children's participation in decision-making and the socio-cultural practices that govern that, rather than

[2] Please note that the information here is hypothetical – it is intended to illustrate a process, not to give information on the substance of the programs used for illustration. For example, they may be sufficient existing research on hypotheses that we state have not been tested before. Again, we are interested in illustrating a process, not in giving substance-area information.

around how to meet basic needs. So, we set this hypothesis aside, and document why we think we should not invest in testing it at the moment.

*Hypothesis 2:*

1. Has this hypothesis been tested, by us or someone else, before? Yes, there is plenty of research out there about the effects of diversifying sources of income on levels of household income. Document the sources of this information, for later use, and set the hypothesis aside.

*Hypothesis 5:*

1. Has this hypothesis been tested, by us or someone else, before? No, after an extensive literature search, we are unable to find any sources of evidence on these questions. So, we proceed to ask other questions about this hypothesis.
2. Is this hypothesis crucial to our work? Yes, we have a number of pieces of work organized around OVC's participation in decision-making: projects, alliances with organizations that work on the same issue, and advocacy work. We are looking to demonstrate that a model we have works, and to lobby the government to implement that model in other parts of the country. This is a unique contribution we can make. So, we proceed to ask other questions about this hypothesis.
3. Is it feasible, given the resources we have, that we will be able to test this hypothesis in a way that will produce rigorous, believable evidence? Yes, we have two different projects into which we can embed different pieces of the research; we have partnered with a university that does work on voice and participation of different societal groups; and we have made a plan to test the hypothesis over time. So, we identify *Hypothesis 5* for testing and proceed to the next step.

***Step 3: Select an appropriate methodology for testing the hypothesis, based on these criteria:***

So, we have decided to test

*Hypothesis 5:*

If OVC participate in decisions that affect their lives, then that will help lift the socio-cultural barriers to their empowerment.

Now we need to figure out how to go about testing this hypothesis, or what the most appropriate and strategic method of testing it would be. Let's ask ourselves a few questions again:

1. What exactly do we need to know? Generally, we need to know whether having OVC participate in decision-making will help lift barriers against them. More specifically, we need to know what exactly these socio-cultural barriers are, how they function within communities, how OVC can participate in decisions that affect them, and how exactly that will help counter these specific barriers. These are research questions we need answered.
2. Why exactly are we testing? Do we need to verify something for ourselves, or do we have an audience we are trying to convince. Let us, for the sake of this example, imagine that we are trying to convince an external audience – such as the government – that this approach will bring positive social change. For this audience, if we do a couple of case studies in areas where we have projects, we may not be able to convince them that the approach we are advocating for will work elsewhere. We will need to show them that the effects we have found in some places are generalizable to other places and communities.
3. What information already exists on our questions? We have done some investigation and research (with the help of a student research intern we had last summer), and found that there is some data on socio-cultural beliefs in Burundi that discriminate against HIV orphans and their participation in community decision-making processes. We have also found that we need to get some data on levels of participation and on particular decisions that arise with children heads of household not traditionally tackled by children in communities.
4. What are appropriate methods to get the information that we are still missing? For example, one is to do focus group discussions with the communities and with the OVC in them. Another is to do key informant interviews with community leaders. Yet another is to do a survey to see whether the particular levels of participation we have found in one community hold true in others as well. Some of these are qualitative, others quantitative methods. The survey will likely be most expensive because we will need technical help designing the questions, testing them, then

sending enumerators to collect the data, then cleaning up the data and analyzing it. The qualitative methods we can carry out via a combination of in-house, staff time and some external help.

5. Which method do we choose? If we only do qualitative, we may not be able to show our audience in a convincing for them way that our approach can be generalized throughout the country. On the other hand, since we have a research partnership with Université Espoir d'Afrique, they will help us with the quantitative survey; what we need to invest is some staff time and a smaller amount of funds for a workshop to organize the research. Based on this, we will take a mixed-methods approach.
6. We document, document, document our thinking in these steps so that we can present it when dealing with the university, the rest of CARE, and of course our audience for the testing, the government.

***Step 4: Test the hypothesis, following the selected methods rigorously***

Since we have a research partner with expertise in methodology, we will discuss the program TOC and strategic decisions with them, but then let them take the lead on carrying out the research itself. They may collect secondary data (and we may need to make some CARE data available to them); they may collect primary data; they may put together interview guides for us to use; they may send some of their students to do research; they may do analysis for us. In the end, we are interested in understanding the findings, with their help, rather than simply having a report from them. Read more about this in Chapter 3.

***Step 5: Analyze what the findings mean for the TOC***

Here, we consider what the testing of the hypothesis tells us about our TOC and future course of action in the program. For example:

- Did we find that there were particular socio-cultural barriers to OVC's empowerment that we did not understand before? If so, how do we refine our pathways to include those?
- Did we find that the low levels of OVC participation in decision-making do not have to do only with community attitudes and practices, but also with information and skills OVC lack? If so, what do include in the pathways – or new pathway – of the TOC?
- Did we find that their participation does not do anything for their empowerment? That is, even when they do participate in decision-making, do the barriers to their empowerment remain in place? If so, what does that imply for revising our pathways and domain in the TOC?

Many more examples can be given, but the point is, we need to understand how the findings from testing the hypothesis bear on the TOC. If we do not go back to the TOC and use what we have learned to improve it and to fine-tune our approaches, then the testing has not done its purpose.

So far, we have gone through all the different steps of testing the TOC, both conceptually and with the example provided by CARE-Burundi. It is key to remember that we test our TOC in order to revise and refine it, and in order to proactively adapt and improve our work over time. The improvement is both in terms of quality (for example, what approaches we take) and in terms of results (the outcomes and impact our work contributes to). Now we turn to ways to check whether our work is producing the results we aim for: tracking changes toward our impact goal.

***PROCESS 2 – Tracking Changes toward the Impact Goal – Some Ideas and Approaches***

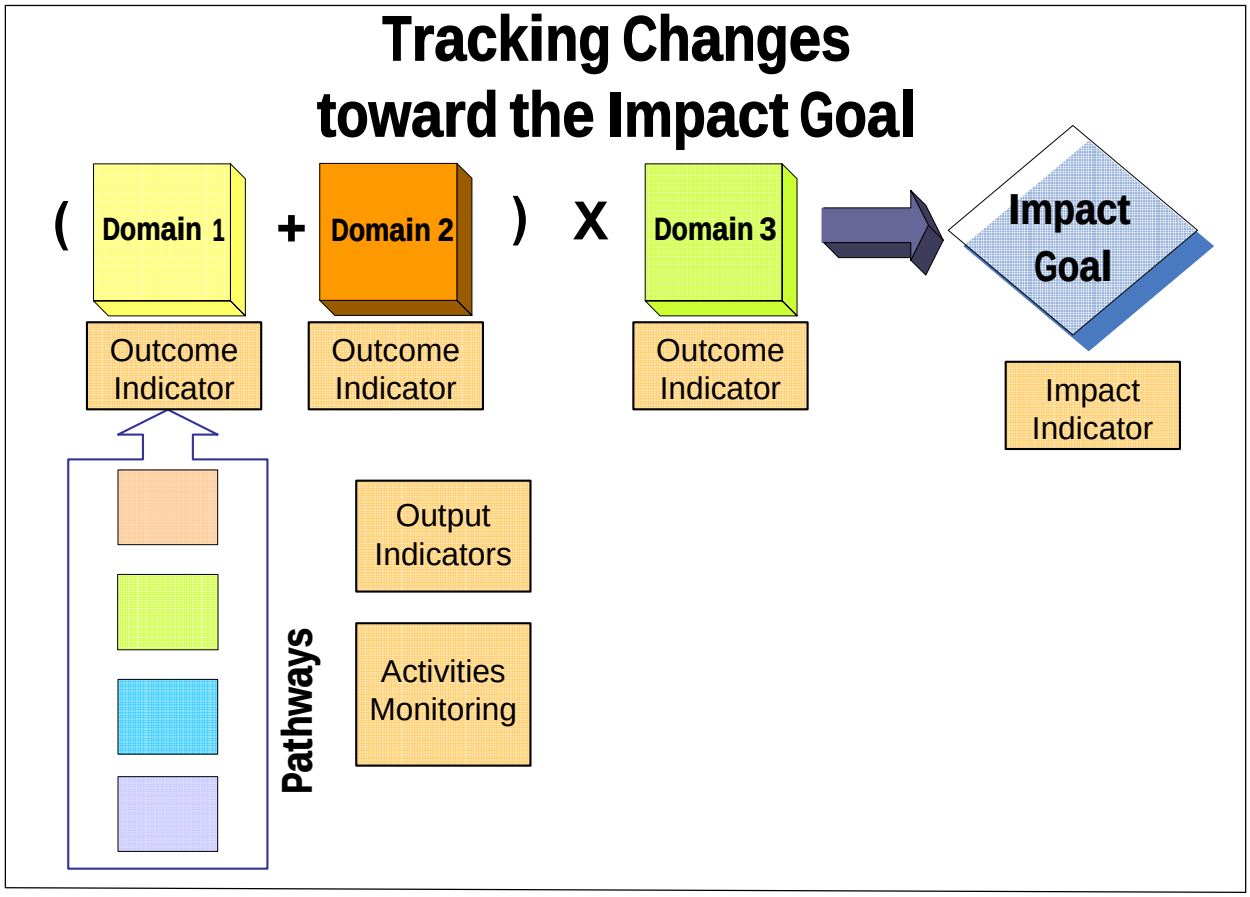
We have a TOC with a 10-15 year impact goal, and along the way we can test hypotheses on different levels of the TOC to validate our thinking, adjust and refine our approaches and strategies, and convince others that models can be scaled up. But how are we going to know whether in the course of our programming we are making progress toward our impact goal?

Tracking changes toward the impact goal allows us to measure our progress toward the social change we envision in the lives of impact groups. This is the second key process, besides testing the TOC, that an

IM system has to support. In the IM system, it is key that the process of tracking changes toward the impact goal be based not on indicator selection or on massive data collection, but rather on analytical thinking about the social changes we want to see take place. So, before we offer step-by-step guidance on how we can approach tracking changes toward our impact goal over time, we want to emphasize critical thinking and analysis. More detail on this is offered in Chapter 3 as well, but here, we should be clear: the first task of tracking changes toward the impact goal is for us, as a program team, to achieve clarity on what deep, sustainable changes we want to see in each domain and impact goal of our TOC. The technical work of assigning indicators, collecting data on them and so on is secondary to the analytical thinking required.

Generally, in order to track progress toward the impact goal, we will want to see outcome-level changes for our domains of change and impact-level changes for our impact goal. Within the pathways of the TOC, we will still have output indicators and activity monitoring. The projects or initiatives that are part of the program can help us track activities and outputs in their M&E plans, much like we have done traditionally. In IM, we are interested in advancing to the levels of outcome and impact. So we will be looking to track outcome-level change with domains and impact-level change with the impact goal.

The diagram below summarizes this idea:




With that in mind, here are some steps to use in tracking changes toward the impact goal. A detailed example of them is provided shortly.

► *Step 1:*
*Identify outcome indicators related to the domains of change*
→ use CARE's Menu of Global MDI+ Indicators

Once we have clarified and stated succinctly what change we want to see in each domain and the impact goal – and only then – we can proceed to selecting an indicator to help us measure if those changes are occurring. To select an indicator, we turn to CARE's global menu of MDI+ indicators. (Refer back to the Introduction for some background on them and for general guidelines on indicator selection, and to the Appendix to this chapter for the menu of indicators.) From the menu, we are seeking one indicator per domain to help us capture outcome-level change. For each domain of change, we look at the right hand side column of the menu, and we select one appropriate outcome indicator.
As you think about only selecting one outcome indicator per domain of change, recall our discussion in the Introduction: with indicators, less is more, and we only need one of them per concept or change we are seeking to measure. We do not need to assign multiple indicators to measure the same change. So for example, we may see four different outcome indicators related to income poverty reduction:

1. % people whose net income generated within target value chains has increased, by sex
2. % households with access to secure land tenure, by sex of the head of household
3. % households that do not rely solely on agriculture for their livelihood, by sex of the head of household
4. % households with capacity to cope with environmental shocks without depleting assets, by sex of the head of household

However, we do not need all four for the same domain of change – we only need one per domain.

Then, we go to Step 2 for the same process but applied to the impact goal:

► *Step 2:*
*Identify impact indicators related to the impact vision*
→ use CARE's Menu of Global MDI+ Indicators

This time, we look for an impact indicator in the middle column of the menu. We are only looking for one impact indicator for the impact goal. As before, you may see more than one indicator that applies to your impact goal. However, it is not necessary to have more than one impact indicator.

Thus far, we have selected one outcome indicator for each domain and one impact indicator for the impact goal. Given that most programs have three to four domains, this leaves us with three to five indicators overall. ***You should aim for 3-5 indicators (one impact and the rest outcome) per program maximum.*** It may be tempting to select many indicators from the menu: but remember, we will need to work on outcome and impact levels here and we should not bite off more than we can chew. Overburdening the IM system with high-level indicators makes it difficult and costly to work with without necessarily adding value to our analyses and learning. Refer again to the Introduction and to Chapter 3 for a more detailed discussion on indicator selection.

Having selected our three to five outcome and impact indicators, we now need to consider how to collect data on them, in Step 3:

► *Step 3:*
*Identify sources of data for all indicators, based on the following criteria:*
→ are there secondary sources of data we can use instead of collecting primary data (e.g., Census, DHS, HDR, etc.)
→ if not, what is the appropriate method of primary data collection?

The key idea in Step 3 is to stick to our rule of thumb (discussed in the Introduction and in Chapter 3): we will not begin primary data collection unless we are absolutely certain that there is no source of secondary data that we can use. If at all possible, we will use data collected by others because usually this is less costly and more reliable. There are many research organizations, institutes, universities, and development agencies that collect and make available all sorts of data related to development topics. For example, we can look at the DHS, UN data, OECD data, the Human Development Report (HDR), and many others. In FY11, the Pi team will update these guidelines to offer specific guidance on sources of data, the strengths and weaknesses of each, and how to use them. The point at the moment is, we will carefully consider sources of data to use before deciding we must collect our own data. Even if their categories do not seem exactly like the ones we would use to collect data, usually we can find categories of data close enough to serve our purposes.

Having found a source of secondary data for each indicator, we need to figure out how often that source makes new data available:

► *Step 4:*
*Specify the time interval for data collection for each indicator, with reference to its source (e.g., when is the Census conducted? When does the DHS come out? Etc.)*

For example, some UN data are released every year. Some DHS surveys are carried out every five years. For each source, we need to know how often it comes out with new data so we can collect those data every time they are available.

Then, each year we will collect data available to us, mainly from secondary sources, keep those data together in a dataset for the program, and periodically analyze to see if over time we are making progress toward the impact goal:

► *Step 5:*
*Establish a baseline for all indicators based on the closest, in time, release of the source of data*

► *Step 6:*
*Every year, collect the value on each indicator from its relevant source*

The gist of Steps 5 and 6 is that once we have sources of data lined up, we will get the most recent data from each (that will serve as our baseline), then revisit them each year, get the most recent data again, and accumulate those data over time. So what our tracking looks like will not be just a single number on an outcome indicator, for example, but rather an accumulation of data points over time.

Finally, we proceed to the single most important step in tracking changes toward our impact goal:

► *Step 7:*
*Every 3, 4 or 5 years, analyze trends based on the collected data and write up analysis*

As we have emphasized time and time again, if we do not do analysis, data collection is pointless. We must analyze and interpret what those data are telling us. And we must document what we have analyzed and learned so we can use it for improving the quality and results of what we do. We need to know whether change has happened from one time period (for example, three years ago) to the next (now). We need to know if the change is in the direction that we expected – positive or negative – or in other words, have things improved or not? And we need to understand what has happened to enable or to hinder change.

Let us consider a practical example of how this works.

***PROCESS 2 – Tracking Changes toward the Impact Goal – An Example from CARE-Ghana***

CARE-Ghana is currently working on developing a natural resource management program which emphasizes control over natural resources by local communities, especially women.

**Issues to be Addressed by CARE-Ghana's FNR Program**

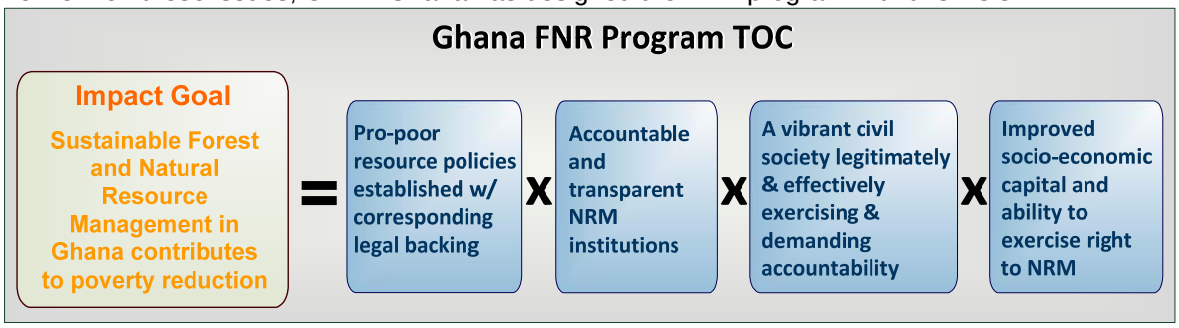
From land use to human rights abuses, Ghana faced a range of problems resulting from mining activities, including:

**Land Use Issues.** Over 80% forest resources (8.2m ha to <1.5m ha) were lost. The extractive sector has not been contributing to poverty reduction and livelihood improvement. The mode of land acquisition included forced evictions, low compensation (and mode of assessment of compensation), resettlement problems, and a non-existent land use plan.

**Pollution Issues.** Wacam's 2009 research indicated that about 250 community rivers in the Tarkwa and Obuasi areas are polluted. Acid mine drainage, cyanide spillage, and seepage into bodies of water destroyed flora and fauna, and degraded the aesthetic value of rivers. In addition, abandoned mine trenches, cyanide containment ponds, and mine rock waste/mine waste disposal polluted fertile lands.

**Human Rights and Quality of Life Impacts.** As a result of the mining operations, more than 60% of the population (a majority of which were poor men and women) lost their sources of income. In response, people protested the loss of livelihoods and lost land and forest. To quell the protests, the government (working in conjunction with the mining companies), turned against its own people. Some suspects were placed in private detention facilities of mining companies; others were arbitrarily arrested. Security agencies, acting on behalf of the mining interests, shot, maimed, and in some instance killed peaceful demonstrators.

To work on these issues, CARE-Ghana has designed the FNR program with this TOC:




While this TOC is still draft and in need of refinement, we will use it to illustrate and emphasize two main points about tracking changes toward an impact goal:

- If we select our indicators strategically, we will need very few of them. This will make our data collection easier and still give us the opportunity to do rich analysis and learning. If we put in the work of conceptually clarifying what underlying change we want to measure, we can find a simple way of measuring it with a single indicator. We do not need many indicators.
- We need to do analysis. If we do not analyze what the data collected on the indicators are telling us, it does not matter how many or how good our indicators are – we will not learn anything and we will not be able to track our progress and improve our program.

In Steps 1 and 2 of tracking changes toward the impact goal, we are looking to strategically select a few outcome and impact indicators out of CARE's global menu. What is a strategic indicator for the FNR Program?

In the global menu of MDI+ indicators, we find the following indicator under Environmental Sustainability: *% local actors with meaningful participation in productive natural resource management at community level, by sex.*

This indicator shows change in more than one domain of the FNR program, as follows:

- Local actors controlling productive natural resources would address their socio-economic status and ability to claim rights to natural resources and their management.
- This community level control would tell us about governance, including decentralization, and the ability of citizens to make decisions (rather than a government selling those decisions to the extractive companies). Applying this indicator to Ghana's impact group of the poor and marginalized would also tell us about pro-poor policies and NRM institutions.
- The disaggregation by sex would tell us about whether women were empowered to meaningfully participate in NRM decisions.
- This indicator illustrates changes across all four domains of the program.
- This indicator also tells us about the levels of the Unifying Framework: social positions (of women, of the poor, and others excluded from meaningful participation) and the enabling environment (of governance and the legal backing for local NRM).

With this indicator identified in Steps 1 and 2 of the process of tracking changes toward the impact goal, we would proceed to research whether there are sources of secondary data we can use (our Step 3), then go back to each of those to collect new data from it when available (Steps 4, 5, and 6). Our preference will be for sources of data we can use, rather than for having to collect data ourselves. We will cast a wide net and carefully evaluate different sources of data before we decide we must collect the data ourselves. And when we do, we will consider a research partnership to bring the expertise we do not have in our organization – see Chapter 3.

Finally, in Step 7, we do analysis. We have seen in this example that, with some analytical thinking before indicator selection, we can strategically select one, relatively simple, indicator that can do multiple duties in our program IM work. However, none of what this indicator reveals above is immediately apparent in the indicator formulation itself. For example, it does not explicitly cite the enabling environment or legal backing, nor does it mention women's empowerment explicitly. Drawing these connections, and explaining what a number collected on this indicator means, is the essence of analysis. Simply collecting data on this indicator cannot replace the analytical thinking that must be done around the change in the domains that the indicator touches upon, the levels of the Unifying Framework it reflects, or the impact areas (Women's Empowerment, Governance, Environmental Sustainability, etc.) that it informs. Without analysis, it means nothing to report to a donor or a partner that 20% of local actors meaningfully participate in NRM at community level, up from 5%. Is this high or low? Is this a significant change or not? Has this change worked the way our TOC hypothesized?

Remember, selecting indicators represents only the initial steps of the process of tracking changes toward the impact goal. An indicator in itself cannot tell us whether we are making progress toward our impact goal. Only the analysis and interpretation of data collected on that indicator can. We have to do the analysis that provides explanations, draws connections, places numbers in context, and tells us what this implies for our programming going forward.

## Chapter 2 – APPENDIX

The appendix to Chapter 2 contains the following materials:

1. Step-by-Step Guidance: Testing the TOC (summary version)
2. Decision-making flowchart for prioritizing hypotheses to test
3. Step-by-Step Guidance: Tracking Changes toward the Impact Goal (summary version)
4. CARE's Menu of Global MDI+ Indicators – FY10 Working Draft
5. Step-by-Step Guidance: Continuous Contextual Analysis (summary version)

CO Impact Measurement System

# *Testing the TOC*

► *Step 1:*

*List all hypotheses that are part of the TOC, and identify their type:*

→ hypotheses related to pathways (within pathways)
→ hypotheses linking pathways to domains
→ hypotheses between domains
→ hypotheses linking the domains to the impact vision

► *Step 2:*

*Prioritize hypotheses to be tested, based on the following criteria:*

→ has the hypothesis been tested before, by CARE or by someone else?
→ is the hypothesis strategically important to CARE's work?
→ is testing the hypothesis feasible given our resources (financial, skills, time, etc.)?
→ document your process of selection of hypotheses to be tested – this becomes part of your arguments and communications to donors, partners, CARE, and others

► *Step 3:*

*Select an appropriate methodology for testing the hypotheses, based on the following criteria:*

→ what do you need to know? What specific knowledge is the testing supposed to produce?
→ what purpose will the knowledge be used for (for example, is this internal validation for CARE? Are you aiming to convince someone else? What does that audience find convincing?)
→ how do we get this information/what needs to be measured?
→ what information/knowledge already exists and how can we use it to answer our questions?
→ what are the possible relevant methods for coming up with the information/knowledge that does not yet exist?
→ what are the resource implications of each methodological choice above (human resources and skills, time, technical, financial)?
→ how do the methods compare?
→ document your process of method selection – this becomes part of your arguments and communications to donors, partners, CARE, and others

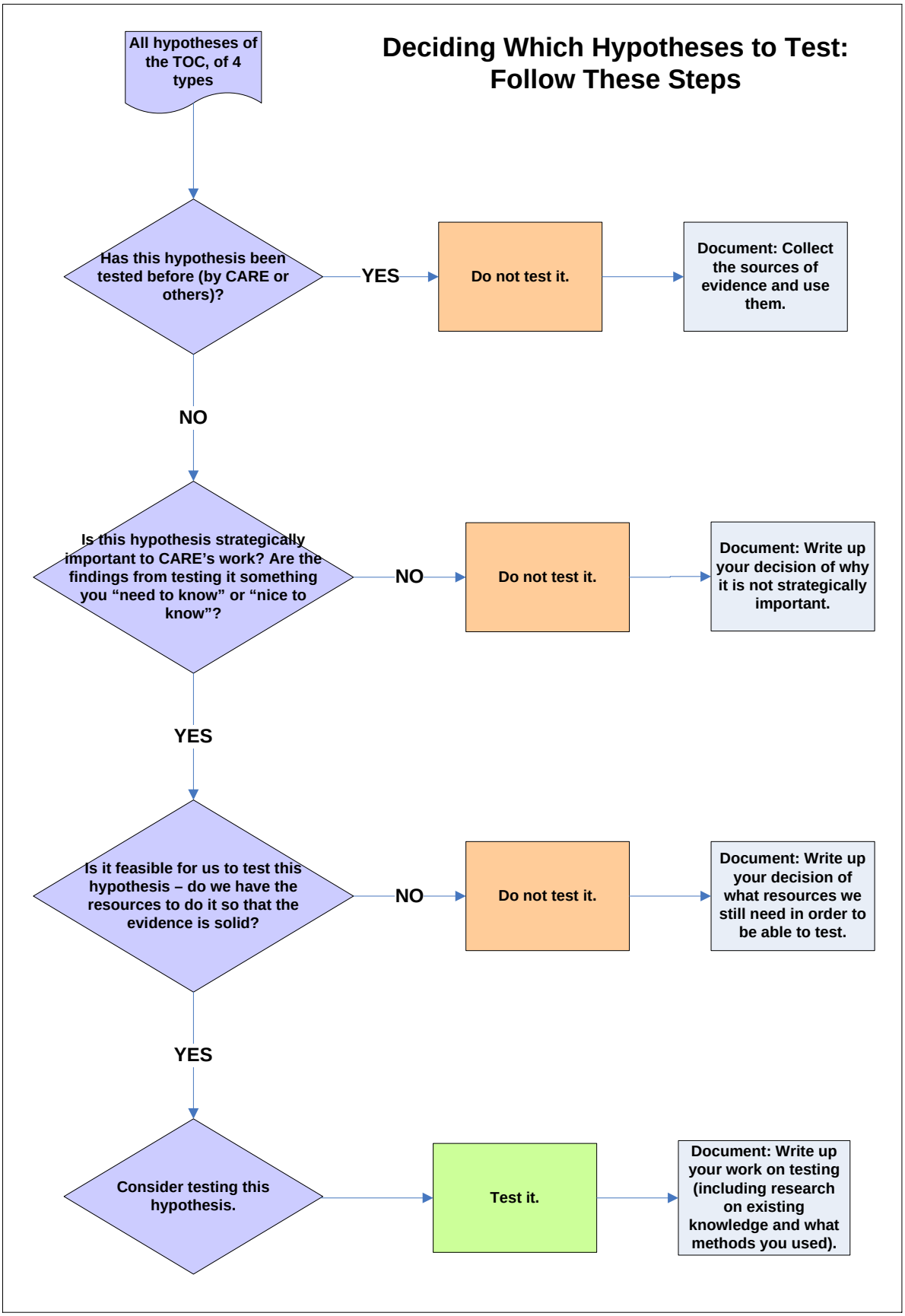
► *Step 4:*

*Test hypotheses, following the selected appropriate method rigorously*

► *Step 5:*

*Write analysis of the TOC, considering the following questions:*

→ what is the role of the hypotheses you have tested within your TOC? Why are they so vital to test and "prove"?
→ what does your test of these hypotheses say about your TOC?
→ what does it say about models you want to show are working and their scaling up?


Deciding Which Hypotheses to Test: Follow These Steps
All hypotheses of the TOC, of 4 types
Has this hypothesis been tested before (by CARE or others)?
YES
Do not test it.
Document: Collect the sources of evidence and use them.
NO
Is this hypothesis strategically important to CARE's work? Are the findings from testing it something you "need to know" or "nice to know"?
NO
Do not test it.
Document: Write up your decision of why it is not strategically important.
YES
Is it feasible for us to test this hypothesis – do we have the resources to do it so that the evidence is solid?
NO
Do not test it.
Document: Write up your decision of what resources we still need in order to be able to test.
YES
Consider testing this hypothesis.
Test it.
Document: Write up your work on testing (including research on existing knowledge and what methods you used).

CO Impact Measurement System

# *Tracking Changes toward the Impact Goal*

► *Step 1:*
*Identify outcome indicators related to the domains of change*
→ use CARE's menu of global MDI+ indicators

► *Step 2:*
*Identify impact indicators related to the impact vision*
→ use CARE's menu of global MDI+ indicators

► *Step 3:*
*Identify sources of data for all indicators, based on the following criteria:*
→ are there secondary sources of data we can use instead of collecting primary data (e.g., Census, DHS, HDR, etc.)
→ if not, what is the appropriate method of primary data collection?

► *Step 4:*
*Specify the time interval for data collection for each indicator, with reference to its source (e.g., when is the Census conducted? When does the DHS come out? Etc.)*

► *Step 5:*
*Establish a baseline for all indicators based on the closest, in time, release of the source of data*

► *Step 6:*
*Every year, collect the value on each indicator from its relevant source*

► *Step 7:*
*Every 3, 4 or 5 years, analyze trends based on the collected data and write up analysis*

# Global Impact Indicator Categories

## With Outcome Indicators

## - working draft FY10- do not distribute outside CARE-

| Category | MDI Impact Indicator | Associated Outcome Indicator |
|---|---|---|
| 1. Income Poverty Reduction | • Proportion of population below $1 (PPP) per day | 1. % people whose net income generated within target value chains has increased, by sex<br>2. % households with access to secure land tenure, by sex of the head of household<br>3. % households that do not rely solely on agriculture for their livelihood, by sex of the head of household<br>4. % households with capacity to cope with environmental shocks without depleting assets, by sex of the head of household |
| 2. Food Security | • Prevalence of underweight children under 5 years of age<br>• Proportion of population below minimum level of dietary energy consumption | 5. % change in dietary diversity, by sex<br>6. % children 0-23 months who are underweight (weight for age), by sex<br>7. % households with food reserves, by sex of the head of household |
| 3. Education | • Net enrolment ratio in primary education<br>• Proportion of pupils starting grade 1 who reach last grade of primary school<br>• Literacy rate of 15-24 year-olds, women and men | 8. Primary school completion rates by formal (gov't funded)/non-formal schools, by sex<br>9. Primary education programs gross enrollment rates by formal (gov't funded)/non-formal schools, by sex<br>10. Student-to-teacher ratios by formal (gov't funded)/non-formal primary schools |
| 4. Women's Empowerment<br>*Mandatory for all programs* | • Ratios of girls to boys in primary, secondary and tertiary education<br>• Share of women in wage employment in the non-agricultural sector<br>• Change in women's self-efficacy | 11. % men and women reporting meaningful participation of women in decision-making at the household level in a domain previously reserved for men (domain:_____)<br>12. % men and women reporting meaningful participation of women in the public sphere (domain: _____)<br>13. % men and women with changed attitudes toward gender-based violence<br>14. % couples making informed joint decisions regarding sexual and reproductive health<br>15. % men and women reporting ability of women to effectively control productive assets<br>16. % women reporting an improvement in their psychosocial wellbeing<br>17. Average number of hours per day spent on house work, and in relation to the duration of the working day, by sex |
| 5. Child Health and Nutrition | • Under-five mortality rate<br>• Infant mortality rate | 18. % children exclusively breast-fed within the first 1 hour after birth and up to age 6 months<br>19. Proportion of 1-year-old children immunized against measles<br>20. % children age 0-23 months who slept under an insecticide-treated bed net in the previous 2 weeks |

| Category | MDI Impact Indicator | Associated Outcome Indicator |
|---|---|---|
| 6. Maternal Health | • Maternal mortality ratio<br>• Proportion of births attended by skilled health personnel either at home or at a health facility<br>• ANC and PNC (at least 1 visit and at least 4 visits) | 21. % women with met need for emergency obstetric care<br>22. % women attending 4 ANC visits at a health facility<br>23. % women reporting satisfaction with the quality of care received |
| 7. Sexual and Reproductive Health | • Contraceptive prevalence rate<br>• Adolescent birth rate<br>• Unmet need for family planning | 24. % people whose need for family planning services is met, by sex<br>25. % people making informed decisions about their contraceptive use, by sex<br>26. % adolescents with access to contraceptive methods, by sex |
| 8. HIV/AIDS | • HIV prevalence among population aged 15-24 years<br>• Condom use at last high-risk sex<br>• Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS | 27. % people with correct knowledge of the 2 major ways of preventing the sexual transmission of HIV/AIDS, by sex<br>28. % adults with more than one partner in the past 12 months reporting the use of a condom during last sexual intercourse, by sex<br>29. % people with advanced HIV infection receiving antiretroviral combination therapy, by sex<br>30. % people utilizing HIV-associated clinical services such as VCT, STI, TB, by sex |
| 9. Environmental Sustainability | • Proportion of population using an improved drinking water source<br>• Proportion of population using an improved sanitation facility | 31. % local actors with meaningful participation in productive natural resource management at community level, by sex<br>32. % of population using an improved drinking water source<br>33. % of population using an improved sanitation facility<br>34. % of population with access to emergency warnings<br>35. % of population using improved fuel sources |
| 10. Governance<br>*Mandatory for all programs* | | *Section under construction – Indicators to be developed by CARE-UK and CARE-USA in FY11* |

CO Impact Measurement System

# *Continuous Contextual Analysis and Trends*

► *Step 1:*

*Identify the aspects of the context to be analyzed, keeping in mind that contextual analysis is NOT the same as UCP/V analysis. In the analysis, include implications for each of your Impact Groups.*

→ regional context, plus global trends as applicable
→ national context
→ local context

For each, → social, political, geo-political, cultural dimension
→ demographic dimension
→ economic context, including NRM
→ environmental context
→ development space context, including donors

► *Step 2:*

*Prepare context analysis matrix, summarizing findings for each of these aspects and identifying key trends (highlights of the analysis) to be watched over time. The analysis should also include considerations of future scenarios/forecasts, as relevant:*

| | Local | National | Regional/Global | CARE CO: Implications for Impact Group |
|---|---|---|---|---|
| Political | | | | |
| Economic | | | | |
| Social | | | | |
| Cultural | | | | |
| Geo-political | | | | |
| Environment | | | | |
| Development | | | | |
| Trends: 1._____; 2._____; 3._____ | | | | |

► *Step 3:*

*Identify critical elements of the contextual analysis and trends applicable to each program*

► *Step 4:*

*Prepare individual context analysis matrix for each program reflecting the critical elements identified for that program*

► *Step 5:*
*Identify trigger factors for the contextual analysis. Triggers of analysis are events or processes that signal to you that an analysis of the context, or an update of your contextual analysis, is needed again:*

→ Internal to CARE

- LRSP process (review or development)
- AOP (review or development)
- Design of new program
- A key or large assessment process related to impact tracking of changes toward impact vision

→ External to CARE

- Major political change
- Major shift of strategic donors (priority, policy, focus)
- Major change in government plans, priorities, etc.
- Etc.

► *Step 6:*
*If any of these triggers of analysis occurs, conduct a contextual analysis, keeping in mind the following criteria:*

→ select an appropriate method for the analysis and apply it rigorously

→ the first time a trigger occurs, do a complete detailed analysis, as defined in the context analysis matrix. The second, third, etc. times a trigger occurs, focus on those areas that have experienced significant change since the last analysis, only updating those cells of the matrixes that need updates.

→ reflect on the trends you have identified for tracking

→ review forecasts/future scenarios and update as needed

→ enter the updates the analysis has produced into your matrix

► *Step 7:*
*Conduct a stakeholders analysis, considering the following questions, and reflect the summary of this stakeholder analysis in cells of the matrix as relevant:*

→ actors, networks, movements

→ agendas

→ resources

► *Step 8:*
*Continue updating the matrix as needed, keeping it a living document:*

→ update cells if a change has occurred

→ enter new trends

→ save a new copy of the matrix every time, instead of changing one and the same copy, so you have a record of changes in the context over time and institutional memory

→ keep the matrix to 1 page so it's user-friendly and used

→ make discussions of the context a regular feature of staff meetings

# III. WORKING WITH DATA: SOME TECHNICAL CONSIDERATIONS

As we have been emphasizing throughout this guidance document, impact measurement is about analytical thinking more than anything else; it's about being reflective, critical, analytical thinkers. To improve our work and its impact, we must focus on what we need to know, what we need to learn, and what questions we need answered. The core of IM is not measurement, or indicator selection, or how to do large surveys or other sorts of data collection. *The core of IM is analytical, critical thinking.* To be sure, there are technical aspects to IM, and using the appropriate indicators with a rigorously applied methodology is important. In FY11, these guidelines will be expanded to provide guidance, standards, and tools for the technical aspects of IM. Again, however, the most important process of IM is to think analytically about what we are doing and what questions need to be answered. Indicators and methods will suggest themselves—but only after we do the required analytical thinking.

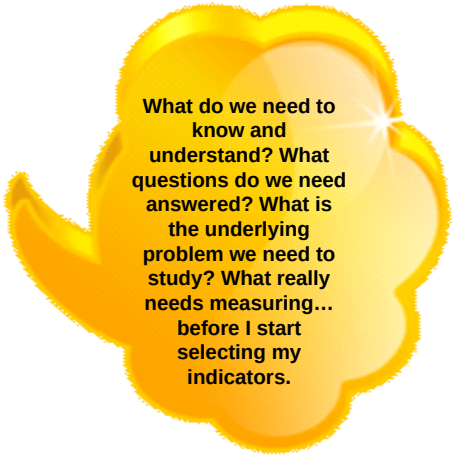
In this chapter we will begin discussing some of the more technical aspects of IM: data collection, data management, data analysis. For the purposes of these guidelines, we will only focus on general approaches, and outline our principles for handling and managing data. In FY11, we will expand these guidelines to include technical guidance on the operationalization of indicators, the selection of appropriate methodologies, the rigorous application of methods, tools for data collection, and approaches to analysis.

While data collection and data management, have a role to play in the IM System, *data analysis* is the most crucial component. We cannot learn directly from data; we can only learn from the analysis and interpretation of those data. Analysis is the critical step between data collection and learning that is often skipped or missed. No data should ever be collected without being analyzed because this is a waste of time, money, and effort, and makes us bad stewards of resources.

What does emphasizing analysis over data collection mean? First, it means leaving the selection of indicators until after the analytical, conceptual thinking has been done. This includes clarifying and specifying what questions we need answered, and knowing exactly what we seek to learn. Once we have done our conceptual homework, the indicators will then suggest themselves. If we have clear questions we seek to answer, the indicators that will help us gather the data to answer them will also be clear.

Second, emphasizing data analysis over data collection means leaving the selection of our methods until *after* the conceptual work has been completed. For example, we often state that we will need to do a large household survey before we have achieved the necessary clarity on what we need to research and what we need to understand. But do we really know that a large household survey would best answer our questions? Or are we choosing it because this is the only method we know, or perhaps because we think that this is what every donor wants to see? A much more effective and convincing approach to our work (to every audience), is to *first* state our questions clearly, *then* select the appropriate method to answer them. Our analytical thinking on what we need to know should suggest our methods, not the other way around.

Another important rule that applies to data collection, data management, and data analysis is: *Keep it simple!* The simplest method of answering a question is always the best—but this doesn't imply that we should ignore standards or methodology. Instead, we must select our methods



carefully and wisely from the various options, then focus on one that we apply rigorously. We must also rigorously follow an established methodology, rather than invent steps as we go along or cobble together steps from different methods. This results in fewer but more meticulously selected indicators, and decreases our massive data collection while increasing our focus and depth of analysis, interpretation, and implications.

The processes of data collection, data management, and data analysis are all part of the larger task of *information management*. Information management entails collecting and managing information from various sources, and distributing it to various audiences. Management entails organizing and controlling the structure, processing and delivery of the information.[3] In sum, information management is about organizing, retrieving, acquiring and maintaining information. Information management, in turn, is part of the larger process of knowledge management and learning. Information management is a crucial component of the IM system overall, and serves as the 'logistical information' for collecting and managing data. However, the information management we do for the purposes of IM does not cover all the information management we need to do in the organization or in a given CO – much like with learning and knowledge management, as discussed in the Introduction. At CARE, we need to do learning and knowledge management for the IM work that we engage in, which, in turn, is part of the larger organizational agenda of learning and knowledge management. Thus, the rules and methods of information and knowledge management apply here, although we will only focus on the three processes of data collection, data management, and data analysis in this chapter.

## Data Collection

### *Primary and Secondary Data Collection*

Once we have selected our indicators, we must determine how we will collect the data on them. One of the first questions to consider is "Do we need to do *primary* or *secondary* data collection?"

In primary data collection, we collect original data ourselves. In secondary data collection, we find and use data (or collect it from various sources) previously collected by someone else. Secondary data collection may entail collecting data that was previously collected externally to CARE (for example, if you use a DHS dataset) or internally to CARE (for example, if you use some data collected by another project prior to your research). In both of these cases, whether data is collected internally (within CARE) or externally (outside of CARE), we are doing secondary data collection.

Primary or secondary data collection? Always choose secondary first! When engaging in data collection, we should do primary if—*and only if*—there is absolutely no secondary source that will meet our needs. Primary data collection is reserved only for those rare occasions when we have exhausted the possibilities of all secondary data collection and data cannot be obtained in any other way.

This is because primary data collection is expensive. Anyone who has done a survey knows the costs associated with developing and testing the survey instrument, training the enumerators for their work with respondents, drawing the sample, entering and cleaning up the data, and so on. When we lack adequate resources to carry out these tasks rigorously, we shouldn't simply omit some of them or do them partially. Instead, we should use valid and reliable data collected by other experts who have had the resources to do the research properly. This is secondary data collection.

**Choose Secondary Data Collection First!**

Another reason to choose secondary over primary data collection is the very specific technical skills and competencies required to plan and carry out a data collection process. Because CARE is not a research organization, few CARE COs have the capacity to do primary data collection without hiring consultants, enumerators, etc. Many COs also struggle with finding qualified consultants who can manage the process properly. The best way to avoid spending scarce resources on a

[3] The discussion here draws on the work of Carnegie Mellon scholars: Barnard, Cyert, March and Simon.

process that is not carried out rigorously is to do secondary data collection. Primary data collection should only be done with much forethought and preparation, and then only when we are certain that we cannot get the data any other way.

To help with finding and accessing data sources for secondary data collection, these guidelines will be expanded in FY11, with a menu of data sources, brief descriptions of what data they contain, and how to access them.

***Selecting the Right Method***

One of the most important steps we need to take before we engage in data collection is to think about the methodology we are going to employ. At CARE, we sometimes have a knee-jerk reaction to doing IM or M&E: we immediately state that we will do a survey. A survey is just one method among many, and often it is not the best method for our needs. As mentioned above, surveys are quite costly in terms human resources, time, and money. Each method should be selected to best suit our needs (what information do we need? who is the audience for our points based on that information? how can we get it most efficiently and effectively? what are the costs associated with that? etc.). We shouldn't use a method simply because "Well, we have always done it that way."

The selection of method should be strategic. One factor to consider when selecting a method is, who is the key audience and what does that audience find convincing? For some audiences, it is perfectly adequate to do a small-scale, informal piece of research. For other audiences, a survey producing quantitative data and analysis may be necessary. Remember that we ourselves are an audience as well: sometimes we may carry out IM or an evaluation because we need some questions answered for ourselves.

Another question to ask when selecting a method is, what resources do we have to carry out the study? Some methods cost more than others to apply; some take more time than others to apply; and still others require very specialized technical skills that may not be easily found in our available consultant base. If we do not have the resources to apply a given method rigorously, it is best that we do not do it at all. A bad piece of research is worse than no research at all.

To help with selecting the appropriate method, these guidelines will be expanded in FY11, with a menu of methods, what they entail in terms of cost, and where to find tools and how to apply them.

***Rigorous Application of Methods***

Once we have chosen the appropriate method, it is crucial to apply it rigorously. We will not be credible if we claim to be following a certain methodology, but then do not do it properly. Sloppy work hurts our reputation, does not give us the convincing evidence we need, and wastes our resources.


All methods – quantitative and qualitative, participatory and non-participatory – have strictly defined rules, standards and procedures. We cannot change, modify, or pick and choose some procedures from one method and some from another, or otherwise deviate from them. Also, it is not enough to simply state that a given context requires that the method be modified. This usually means that the method was poorly selected and a more appropriate one should be found. We cannot change a method's standards and procedures to fit our context.

***Research Partnerships***

In IM work, research partnerships are extremely useful. They can add credibility to our IM and evaluations. They can add a fresh look and the richness of different perspectives on our work. They can bring us technical research expertise that we lack within the organization. They afford us a way to be rigorous in our IM work even when we do not have the research expertise ourselves. Because we do not have, and often do not need to develop, the technical expertise to apply all research methods rigorously,

we should consider research partnerships. The key, however, is to identify the right research partner. Some factors to consider here include:

1. Have we clearly identified the questions we need to answer in the study? It is CARE's job – not the job of the research partner (or consultant) – to identify the questions that CARE wants answered. If we do not, we will repeat the experience of paying for research that does not tell us what we needed to know.

2. Have we identified the general methodologies to be used? Many research institutions have a methodological focus: some only do large-scale quantitative surveys; others only do participatory methodologies; etc. We need to know whom we are approaching and how their methodological expertise will benefit us. Even if we do not know how to implement a particular method, we should know if that is the method that they emphasize and use.

3. What are our research partnership needs? Think strategically about the partnership; it will save us time and resources. There are different partnership needs—sometimes we need a renowned name in a given field to appear on the study next to CARE's name. At other times, it is quite sufficient to recruit a few graduate students with research skills from a university to work with us for a summer.

These are some basic issues we should consider when thinking about data collection and the appropriate methods to answer our questions. To be effective, credible, and rigorous in our work requires us to make informed strategic choices of indicators, of methods, and research partners.

**Data Management**
Data management requires us to follow policies and specific consistent practices that safeguard, control, organize, and deliver our datasets as needed. The underlying notion here is that data are a valuable resource that needs stewardship and management, just like our other resources. The following are some basic principles of data management that our experience has highlighted.

***Ownership of Data: CARE-Sponsored Studies = CARE-Owned Data***
When we do a study (e.g., a baseline, an evaluation, etc.), the data collected and produced in primary data collection are products and property of CARE. We have paid for those data to be collected and organized—*we own them*. This is true even if we hire a consultant or partner with a research institution for the work. We may give others access to the data and allow them to use the data. We may even later hire or partner with others to work on the data for a different study or purpose. But we always retain ownership of our data. Therefore, if we are working with a consultant or partner, we should always demand the raw data as a deliverable. Similarly, if we are putting the dataset together ourselves, we should safeguard and store it properly. It is not good practice to have reports for which we cannot show the raw data.

***Data Accessibility and Good Data Management***
Making data accessible to multiple users is part of good data management. Data that are not accessible to people are not properly managed. Storing a dataset on a single staff member's computer is not good data management because it does not make the data accessible to many people who may need it. Similarly, if that computer crashes or is stolen, the data can be lost forever. Consistent file naming protocols should also be used. Inconsistent file naming protocols make it impossible for users to locate the right data files. An example of a proper file naming protocol is one in which the file contains the project name, the date of data collection, and the source of the data (location if primary data collection, source if secondary data collection). This approach will enable people to locate the data file and understand what it contains.

A more systematic explanation, and specific detailed guidance, for data management will be provided in FY11, as the next iteration of these guidelines. For the moment, we need to be aware that good data management is key to sound impact measurement.

**Data Analysis**

The topic of data analysis is quite vast and quickly becomes highly technical. There are numerous guides and textbooks on how to do data analysis according to a given method. The purpose of these guidelines is not to reinvent or rewrite these data analysis guides. Nor is it to give overviews of data analysis methods, simplify them or make them accessible. Too often, when we start simplifying and modifying analysis methods to make them accessible, we lose rigor inherent to the method and end up doing work that is below technical standards. This section discusses the key components of quality data analysis.

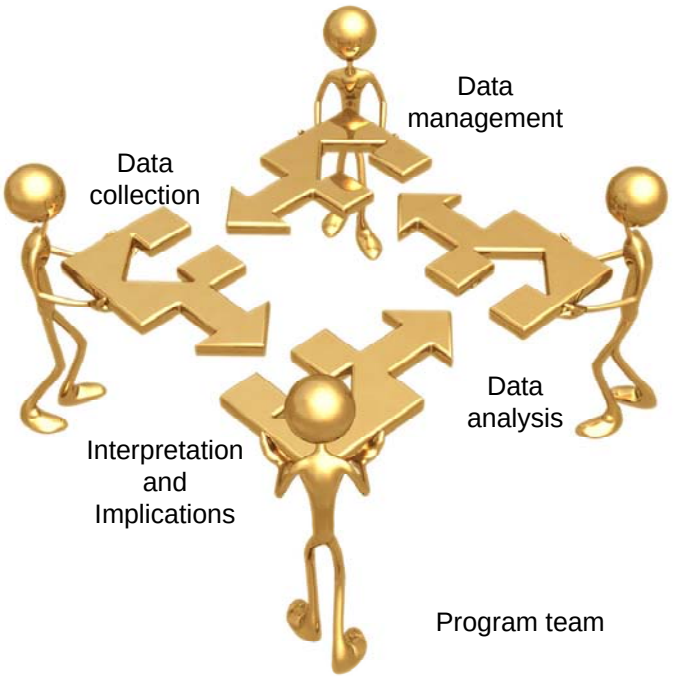
***Remember to do the Analysis***

Simply put, the most important thing to do when we approach data analysis is to make sure it gets done. Sometimes we collect massive amounts of data which then go unanalyzed and unused. This is much like buying the ingredients and then not cooking the meal. It is not a good use of our resources.

***Involve the Right People in the Analysis***

Take a moment to think about who should be involved in the analysis. It is untenable to think that the one M&E officer in the CO will do all analyses for all studies. But it is equally untenable to think that all staff will always be involved in all analyses. It is bad practice to hire a consultant and leave everything up to them, without the proper amount and specific input and guidance by CARE. At the same time, it is unrealistic to assume that all staff can learn the intricate methods of data analysis, whether statistical or qualitative. So how do we determine who should be involved?


Traditional M&E




Impact Measurement

The decision of whom to involve becomes easier if you think about analysis as having three stages or steps:

1. *Findings.* Findings are the facts or information we uncover by doing the research. We have not yet figured out what they mean for us; we are simply stating what we found. For instance, we may

have found that the percent of people with meaningful participation in natural resource management at the community level in Ghana has increased from 5% to 20% over the last five years. Or, we may have found that the patterns of economic exploitation in Bangladesh involve predatory lending at exploitative rates by local community money lenders. This work is best left to a qualified consultant or research partner who can apply a method rigorously and write up the findings.

2. *Interpretations.* Interpretations involve stating what the findings actually *mean*, not just describing what they are. For example, we apply our knowledge of the context, the situation of the impact group, the legal framework in the country, the underlying causes of poverty, and so on. Here, we state whether this increase in people's participation in natural resource management is high or low; sufficiently large or not; as expected given where we started from; etc. This work needs to be done by program and project staff in the CO who are familiar with the programming, the communities, the country context, and CARE's other work. These staff can best interpret whether the findings make sense, whether there is anything unexpected among the findings, and what the findings mean. This work should never be left to a consultant or research partner without CARE's participation.

3. *Implications.* Implications involve stating what the interpreted findings mean for our work, both looking back and going forward. How do we use them to modify our TOC? How do we use them in new program design work? What do the interpreted findings imply for what we have accomplished, or not, and for how we need to reflect that in our programming in the future? This work is the sole responsibility of CARE staff. No consultant or research partner can do it for us. We may take their recommendations. We may also invite the community or impact group, other partners, or other stakeholders to participate in drawing these implications. But it is CARE staff who should be responsible for leading this process.

***Complete the Process of Analysis***

Finally, we should be sure to emphasize that the process of analysis needs to be completed with all three steps above. If we have only listed the findings, we are not done. No report or communication we produce—whether for a donor, the community, a partner, or anyone else—is complete without the interpretation of findings and a statement of what implications we see. Similarly, the internal process of completing a baseline, an evaluation, or any other research is not complete until the appropriate CARE staff interpret the findings and draw their implications. If we want to learn anything, rather than simply compile large amounts of data, we need to carry the analysis consistently through all its stages.

These are some of the basic principles of data collection, management, and analysis that will help us manage these processes. While more technical assistance and guidance on each will follow later, it is important that we all think about them now, as the fundamental principles of solid IM work.

# IV. CONCLUSION: DEVELOPING IM GUIDANCE IN FY11 AND BEYOND

[insert text from Maliha on process for developing the FY11 version of this guidance – PENDING]