

# RURAL

# 21

The International Journal for Rural Development

1 | 2018  
VOLUME 52

## MEASURING IMPACT

Why? How? For whom?



### IMPACT EVALUATIONS

Learning from what works  
and what does not

### LAND DEGRADATION NEUTRALITY

New impetus for the global  
agenda on land degradation

### ORPHAN CROPS

Value chain development for  
better food and nutrition security

## DEAR READER,

Since the adoption of the Paris Declaration on Aid Effectiveness in 2005 at the latest, there have been more calls for a reform of impact measuring in development interventions. One of the reasons for this was the conflicting results that had emerged at various levels. For although the development co-operation organisations had given their programmes and projects positive appraisals in most cases, no clear conclusions could be drawn from the analysis of macroeconomic variables regarding any links between Official Development Assistance and levels of prosperity in developing countries.

Although a number of attempts have been made to explain this so-called micro-macro paradox in development co-operation, the obvious weaknesses in impact measurement in its then existing form could not be ignored: isolated measures of development organisations that were continued at the level of evaluation, a “northern” perspective that did not consider capacity building among stakeholders in the South and methodological shortcomings such as insufficient data establishment and neglecting the issue of what it would really be like for the beneficiaries if the intervention had not been performed. One of the results of such criticism was the launch of the International Initiative for Impact Evaluation (3ie) in 2008. Its aim is to promote independent, methodologically sophisticated and meaningful evaluations. Experience and knowledge gathered so far in the course of 3ie is also reflected in several contributions for this edition.

Our authors describe which impact measurement methods have proved to be useful during the last few years and what their strengths and their limits are. They take a

discerning look at randomised controlled trials, a method that is well-established in the field of medicine and was regarded as an ideal solution for a long time. They demonstrate why participative approaches are so valuable both with regard to learning among evaluators themselves and to taking the multifunctionality of evaluations into account. They show how digital data acquisition solutions can facilitate the work of evaluators and why simply combining conventional methods is not enough when it comes to complex evaluations. And they give a clear impression of how difficult it still is to select suitable indicators and achieve a balance between considering individual conditions and the desire for standardisation and a maximum of comparability in practical results monitoring – despite all the insights gained over the last few decades.

On average, 400,000 US dollars and a period of three years is needed to measure the effect of a development intervention, the International Initiative for Impact Evaluation concludes from its own activities. These are valuable resources that need to be used carefully if the top priority is to be what the authors of our introductory contribution suggest: that impact evaluations first of all benefit the poor.

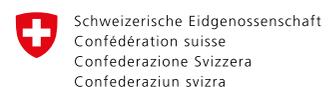
We wish you inspiring reading.



Sincerely yours,

*Silvia Richter*

### Partner institutions of Rural 21



Direction du développement  
et de la coopération DDC



Photo: Simon B. Opladen

**NEWS**

- 04 Is Europe’s agricultural policy fit for development?
- 05.1 Water for all
- 05.2 Development organisations warn of debt crisis



Photo: Malte Lech/DEval

**FOCUS**

- 06 Improving development policies with impact evaluations
- 09 Randomised controlled trials and rural development – an abundance of opportunities
- 12 Randomised controlled trials – the gold standard?

- 15 Learning from participatory evaluations
- 18 Making people visible
- 21 Impact assessment in complex evaluations  
A call for the systematic integration of multiple methods
- 24 More than plug-and-play  
Digital solutions for better monitoring & evaluation
- 26 The indicator challenge
- 29 Using standard indicators – opportunities, challenges and risks
- 30 Corporate-level impact measurement  
IFAD’s experience
- 32 A brief glossary of impact evaluation terms



Photo: WWPics/Redux/laif

**INTERNATIONAL PLATFORM**

- 33 Land degradation neutrality  
A new impetus for addressing the degradation of land and soils
- 36 Measuring land degradation needs to be done from the ground up
- 38 A small effort and a big impact  
From pedestrian to tractorable bridges – experience from Nepal and Laos



Photo: Georgina Smith/CIAT

**SCIENTIFIC WORLD**

- 40 Remembering forgotten crops – developing new value chains

## IS EUROPE'S AGRICULTURAL POLICY FIT FOR DEVELOPMENT?

In late February 2018, members of the European Parliament's Committee on Development analysed how to improve policy coherence for development when reforming the Common Agricultural Policy.

Ever since its inception in 1962, the European Union's Common Agricultural Policy (CAP) has seen constant change. In the 1980s European agriculture had to cope with milk lakes and butter mountains. Government money was spent to sell large surpluses, which disturbed trade. Nowadays, there are no export subsidies, and with the exception of arms, the poorest countries may export anything they want to the EU without duties being charged. But whether this really makes the CAP, whose next reform is scheduled for 2020, "fit for development", was discussed in the European Parliament's Committee on Development (DEVE) in Brussels, Belgium, in late February 2018.

The production approach was still at the core of the new CAP, criticised Olivier de Schutter, Deputy Chairman of the International Panel of Experts on Sustainable Food Systems (IPS-Food), in Brussels. Adama Diallo, dairy herdsman and Chairman of the National Union of Smallholders and small dairies in Burkina Faso (UMPL-B), explained what this meant for the producers in the South. With nine million dairy cattle, the sector accounts for 17 per cent of the country's gross income. Since the 1990s, numerous micro-dairies have developed that take in between ten and 2,000 litres a day for processing. Rising demand for dairy products is also attractive for international corporations, which export the milk powder to Burkina Faso, according to Diallo, who maintains that with just five per cent duty being charged, local dairy producers are not sufficiently protected. Whereas local milk costs around 500 CFA Francs a litre, the same quantity of milk made of skimmed milk powder and palm oil additive can be bought for just 250 CFA Francs. This dumping price was preventing the development of social structures and Burkina Faso establishing a value chain of its own, and it was making the country dependent on food imports, Diallo told the committee.

### CAUSE-AND-EFFECT PRINCIPLE MORE COMPLICATED THAN ASSUMED

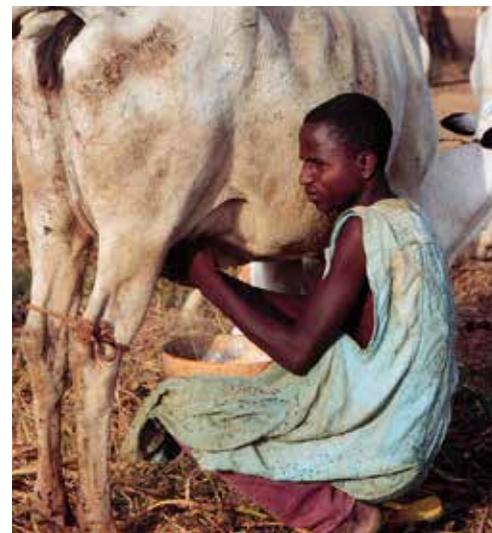
It was not at all easy to circumscribe the negative impacts in the developing countries to trace them back causally to the EU Common Agricultural Policy, according to Maria Blan-

co, an agricultural economist from Madrid, Spain. Blanco examined the latest CAP financing period with regard to such impacts. In the context of world-wide agricultural trade, the CAP had fewer effects on prices and quantities than were commonly ascribed to it. Blanco nevertheless maintained that there were elements that ought to be changed. Unfortunately, the Member States were once again opting more strongly for coupled payments, which was keeping production at a constant level and putting pressure on prices for farmers within and outside the EU.

Nevertheless, this is not a growth guarantee, either. For, according to Blanco, without payments coupled to production, less protein feed is produced, and the demand for feed has to be covered by imports, putting pressure on the natural resources in the region's countries. Therefore, Blanco argues, it has to be carefully assessed which products are supported by coupled payments in which regions and for how long. She maintains that the global treaties of Agenda 2030 and the Paris climate agreement provide an umbrella for a coherent development policy in which the CAP would be one of a number of elements, and concludes that a sector approach is more important than just focusing on the CAP itself.

### A CALL FOR DECENTRALISED, REGIONAL FOOD SECURITY

Olivier de Schutter drew distinctions between several groups of poor people and people suffering from hunger. The major share of them, around 300 million, were living in the slums of the large cities and were producing virtually no food. Local governments could only provide these people with imports. The next largest group is the smallholders, whose production potential, de Schutter claims, has not yet been fully exploited. Smallholders like Diallo produce for the local market, but they have hardly any opportunities to gain access to the domestic market as a whole. In many countries, they compete with the smallest group of farmers who are integrated in the value chain and mainly grow cash crops for exporting. Forty-five per cent of soy cultivated world-wide is exclusively for export, and maize and sugar show similar figures.



Gaining market access is difficult for most smallholder dairy farmers.

Photo: ILRI/Dave Elsworth

While international trade was resulting in those crops being cultivated that grew best in certain countries, de Schutter maintained that there could be no mention of a balance between countries with a food surplus and those with a food deficit. He explained that agricultural produce went wherever purchasing power was greatest. Against the background of climate change, decentralised and regional food security was the best option.

Agricultural economist Harald von Witzke explained that agricultural production and dietary habits in the EU required around 30 million hectares of land for food and feed in other countries. Witzke believes that extensifying agriculture through organic farming is the wrong approach. If Europe was 100 per cent organically farm-managed, virtual area import would rise to 70 million hectares. Von Witzke recommended raising productivity in the total existing area under cultivation and making use of available innovations in pesticides and herbicides, genetic engineering and digitalisation as well as modern seed and fertiliser, both in Europe and in the developing countries. The precondition for this was free access to resources. It had become clear that among the developing countries and regions, those with good governance, market economies and skilled labour were making the most progress.

**Roland Krieg**

## DEVELOPMENT ORGANISATIONS WARN OF DEBT CRISIS

Germany's debt relief alliance "erlassjahr.de" and the catholic relief organisation Misereor are drawing attention to dramatic developments in government debt among the poor countries of the Global South. They called on the G20 Ministers of Finance, who met in Buenos Aires, Argentina, in mid-March, to immediately create a debt relief option in order to fend off the catastrophic implications of a debt crisis for the poorest and most vulnerable people in the Global South. In the run-up to the meeting of ministers, the organisations presented their "Schuldenreport 2018" (2018 debt report), according to which 119 developing countries and emerging economies are critically indebted, while the debt situation has further worsened in 87 of them and 13 countries have had to suspend their debt re-

payments. Politically unstable countries such as Burundi as well as countries whose economy is heavily dependent on exports and that are suffering from the collapse of prices on the commodities market, like Angola or the Republic of Congo, are affected in particular, the report states.

The organisations speak of worrying developments, reporting that people are being driven more and more into poverty, while a major portion of budgets goes into debts servicing instead of being made available to social services such as health and education. Growing unemployment and social tension, rising migration levels and instability in whole societies are the consequences, they demonstrate, and warn that continuously rising government

debt is thus also leading to more social and political instability world-wide.

The two organisations argue that while the G20 have opted for promoting private investment in African countries with initiatives such as the Compact with Africa, they are completely ignoring the corresponding growing risk for the debt sustainability of these countries, which they claim is totally irresponsible. In addition, they call for debt relief options for regionally or thematically circumscribed groups of countries. A debt relief option of this kind would, for example, be conceivable for Caribbean island nations that have been especially hard hit by natural disasters or the impact of climate change.

**Silvia Richter**

## WATER FOR ALL

To mark World Water Day on the 22<sup>nd</sup> March, a wide range of organisations called for a fundamental shift in handling the natural resource of water. Not without reason, for already, 40 per cent of the world's people are affected by water scarcity, while a further 700 million people are at risk of being displaced by intense water scarcity by 2030. More than two billion people are compelled to drink unsafe water and over 4.5 billion people don't have safely managed sanitation services. Women and girls suffer disproportionately when water and sanitation are lacking, affecting health and often restricting work and education opportunities.

Wherever water is scarce and has to be transported over long distances, people above all use this precious good for drinking and cooking purposes. Washing regularly is often neglected. However, such a lack of hygiene encourages the transmittance of diseases. One example is the highly infectious bacterial eye inflammation trachoma. World-wide 19 million people have already had their sight impaired or even gone blind through trachoma infection, reports "Christoffel-Blindenmission" (CBM). Washing one's face regularly can be a protection against infection and hence also blindness. But particularly in the world's poor regions, there is a lack of access to clean water – and of appropriate sanitation and knowledge of correct hygienic behaviour. The protestant relief organisation "Brot für die Welt" drew attention



Washing one's face regularly can protect against trachoma infection.

Photo: CBM

to what water scarcity meant for smallholders in particular, pointing out that the countries in the northern hemisphere were contributing considerably to wasting and exhausting valuable freshwater resources through their agricultural imports. At the International Water Forum in Brasilia, Brazil, in mid-March, held under the motto "Water for all", the organisation called for making access to water fairer internationally in order to ensure that local populations were supplied with drinking water, and hence food, in the long term.

"The ecosystems on which life itself is based – our food security, energy sustainability, public health, jobs, cities – are all at risk because of how water is managed today," said

World Bank Group President Jim Yong Kim. In April 2016, the President and the then United Nations Secretary-General Ban Ki-moon convened a High Level Panel on Water (HLPW). The core focus of the Panel was the commitment to ensure availability and sustainable management of water and sanitation for all, Sustainable Development Goal (SDG) 6, as well as to contribute to the achievement of the other SDGs that rely on the development and management of water resources. With their now published report "Making Every Drop Count: An Agenda for Water Action", the panel members call for policies that will allow for at least a doubling of water infrastructure investment in the next five years.

**Silvia Richter**

# IMPROVING DEVELOPMENT POLICIES WITH IMPACT EVALUATIONS

The debate about the impact of development co-operation is as old as development co-operation itself. The debate about the relevance of impact evaluations is a more recent phenomenon. Wouldn't it be better to make use of the already scarce resources in projects instead of spending them on these costly evaluations? Our authors demonstrate when impact evaluations make sense – and when they don't – and why we can't really do without them.

By Bartłomiej Kudrzycki and Isabel Günther

Over the last 25 years, the share of the world population that lives in destitute poverty has dropped from 35 per cent to 10 per cent, and the share of people who are undernourished has fallen from 19 per cent to 11 per cent. These numbers hint at the progress we have made towards eradicating global poverty. Policy-makers, government officials and development practitioners can certainly take some credit for these improvements in people's lives – but how much? Have projects designed, financed and implemented by various organisations contributed to this success, and to what extent?

Most importantly, despite past collective success, what can be done to do even better going forward? About 700 million people still live on less than 1.90 US dollars purchasing power parity a day, and about 800 million are still undernourished. These are unacceptably high numbers.

By bringing rigorous analysis to empirical data, impact evaluations allow to measure the effect of development interventions and to generate knowledge about how a programme works and how its design and results can be improved. Impact evaluations are thus primarily a tool for learning and improving development interventions. They are an important part of a broader agenda of evidence-based policy-making – in contrast to ideologically, emotionally or politically driven policies. For example, impact evaluations have revealed that farmers in developing countries are less constrained by their access to credit than once thought. Instead, a lack of risk coverage or psychological biases appear to be more likely barriers to farmers investing in new crops and innovative technology. Such insights are of great value to policy-makers looking for effective measures to assist farmers in adopting new technologies.

Besides providing insights for policy-makers and development practitioners, impact evaluations first of all benefit the poor. While critics argue that experimentation on the poor is

unethical, it is also evidently unethical to intervene in the lives of the poor without understanding the changes, intended or unintended, that these interventions are likely to bring about.

## IMPACT EVALUATIONS – A NEW BUZZWORD?

Thanks to technological progress in data collection and the ever-increasing availability of data as well as the creation of various institutions promoting impact evaluations, the number of them being conducted has risen rapidly over the last two decades (see Figure on page 8). While in 2000, the International Initiative for Impact Evaluation (3ie) recorded less than 40 new impact evaluations related to development and poverty, by 2012, the impact evaluation repository of 3ie was publishing over 400 impact evaluations a year. Whereas for many years, impact evaluations were focusing on health questions, the number of impact evaluations has been steadily increasing in other sectors since 2006, in particular in agriculture and nutrition.

“ *Impact evaluations first of all benefit the poor.* ”

At the same time, impact evaluation has become something of a buzzword in development co-operation. Major organisations are creating entire funds and policy priorities in their name, while many practitioners are left in the dark about what impact evaluations actually are and how they are used. Several large development agencies have therefore released primers and guidance documents to address this disconnect.

The lack of clarity around the concept, combined with the high cost of impact evaluations in terms of both time and money, have result-

ed in considerable backlash, even resentment, towards impact evaluations – especially in their most famous (or notorious) form, randomised controlled trials (RCTs).

In this issue, various authors hope to clarify and elucidate what impact evaluations are and when they are effective tools for learning: we believe they have become an indispensable tool for measuring and improving the impact of projects and policies on decreasing poverty and – what is equally important – setting precise and realistic aspirations for the future.

## WHAT IS AN IMPACT EVALUATION?

For many years, the development community – including the Development Assistance Committee of the Organisation for Economic Co-operation and Development (OECD-DAC) in its Criteria for Evaluating Development Assistance – used the term “impact” to refer to the final level of the causal theory of change, or log frame. This definition has been replaced in recent years, and impact evaluations are now seen as “an objective assessment of the change that can be directly attributed to a project, programme or policy”. This could be the impact of an information campaign (about the importance of crop rotation) on farmer output, or it could be the effect of introducing rainfall insurance on a farmer's choice of crops. These changes are the impact: the difference in people's lives (farmers' output or choice of crops) with and without the intervention, measured after the intervention (information campaign or rainfall insurance) has taken place.

To assess the impact of a project or policy one needs to know what would have happened to the population in its absence. This is called the counterfactual, which is a crucial component of any rigorous impact evaluation, and which can be estimated using a variety of statistical methods. In contrast to monitoring, the use of the counterfactual methods lets policy-makers



In order to measure the impact of a crop insurance scheme, one needs to know how the farmer would have fared without it.

Photos: FAO/Luis Tato (left), FAO/Lou Dematteis (right)

and other stakeholders establish the causal effects of their programmes and policies.

For example, an impact evaluation might assess the impact of a programme that aims to improve farmer crop yields by offering farmers rainfall insurance. To estimate this impact, one needs to compare the outcomes of farmers who receive rainfall insurance to the hypothetical situation in which the same farmers were not insured. Studies have found that insured farmers grow riskier crops with higher yields. Thus, impact evaluations establish the direct connection between projects or policies and measurable, observable changes in people's lives.

While the term impact evaluation comprises a wide range of methodologies, one of them has garnered the lion's share of funding, attention, and criticism in the development community: the randomised controlled trial (RCT; see also articles on pages 9–11 and 12–14). RCTs are the most well-known form of impact evaluation, but it is very important to note that there are many other methods of constructing a counterfactual to estimate what would have happened to the target population in the absence of the project or policy without resorting to randomly allocating the target group to a control and treatment group.

### MONITORING – USEFUL, IF INTERPRETED CAREFULLY

To better understand what impact evaluations are, it also makes sense to clarify what they are not. Monitoring is a common, yet non-rigorous method of estimating programme effects and is hence prone to errors. Only measuring the changes of outcomes for the population before and after a development programme, there is no way of knowing if the outcome would have remained the same in the absence of the programme. For instance, a monitoring system can observe that the nutrition of a village population improves after everyone in the village has received a crop storage container. However, unless all competing explanations can be eliminated – e.g. changes in agricultural productivity, construction of a new well, changes in income, or the presence of deadly diseases – we cannot be sure that the impact is indeed a result of the intervention.

Monitoring data is nevertheless often used in development work thanks to its ease and low cost for reporting and project evaluations. Monitoring is useful when the focus is on operation, implementation, or service delivery. However, when misinterpreted as evidence of a causal relationship between a development intervention of a programme and poverty re-

duction, conclusions drawn from studies solely using monitoring data can lead to ineffective or even harmful policies – and in most cases to a waste of public resources.

### WHEN ARE IMPACT EVALUATIONS USEFUL – AND WHEN NOT?

Impact evaluations are a tool for policy-makers and development practitioners to improve development outcomes based on evidence. The findings of impact evaluations can help organisations to decide whether to scale up projects with proven positive impacts or discontinue interventions lacking in effectiveness. Impact evaluations can also identify the specific point – of the theory of change – at which policies don't work as planned. For instance, the Agricultural Technology Adoption Initiative (ATAI) shows that index-based weather insurance is very effective when taken up, but that at market premiums take-up is very low (6–18 per cent) – it is at the point of take-up, not after, that rainfall-index insurance programmes seem to run aground. Impact evaluations can help to design development programmes by comparing different interventions with regard to their effectiveness. For example, poor Kenyans were offered a variety of ways to encourage saving – text message reminders, match-

ing ten or twenty per cent of savings before or after the savings period, and a simple, fake gold coin with a number for each week of the experiment that served as a physical reminder of savings. The intervention that helped farmers save the most by far was, remarkably, the gold coin.

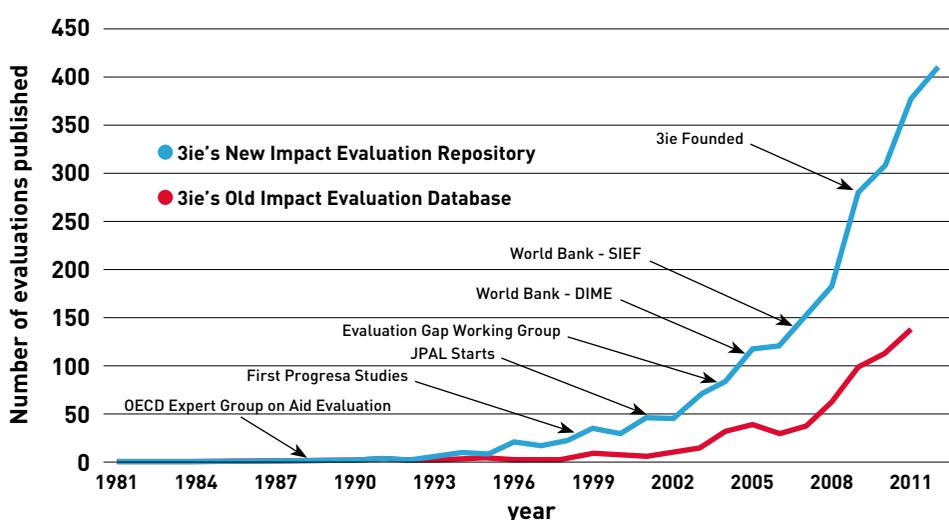
The two main drawbacks of impact evaluations are their high monetary costs and the time required for the results to come back. From the beginning of implementation to the results, an impact evaluation generally takes two years to complete, while many take longer. Average 3ie-supported studies cost 400,000 US dollars and last three years.

Hence, not all projects and programmes of an organisation should be evaluated with regard to their impact; only those where the learning potential is the highest. The project should be strategically or operationally relevant for the organisation and innovative in the sense that evidence on whether it works is needed because impact evaluations on the planned intervention are non-existent. For example, the impact of micro-credit, rainfall insurance and better price information on farmers' livelihoods have already been extensively studied. However, unlike clinical trials in medicine, the findings from impact evaluations (and RCTs) in agricultural development do not easily translate from one context to another. Rather than just providing estimates of the effects of cookie-cutter interventions, impact evaluations should hence be designed in a way to offer the opportunity to learn how context and intervention interact. For any individual study, there is little certainty that the findings will replicate in another context.

Once the number of studies run in different contexts reach a critical mass, however, impact evaluations can inform policy-makers and donor organisations whether they are following the best strategy for achieving a certain development goal; and be used for global policy-making and best practices. Systematic reviews use a structured approach to summarise the results of many impact evaluations from a particular sector or region, and give reliable indication about the success of a certain type of intervention – and are particularly useful for policy-makers and practitioners. For example, the recent evidence maps of 3ie on agricultural innovation and agricultural risks are great starting points for policy-makers working on rural development and agriculture.

Neither is it possible to analyse the impact of all types of development interventions. In oth-

Impact evaluations published by year, 1981–2012



SIEF = Strategic Impact Evaluation Fund; DIME = Development Impact Evaluation; JPAL = Jameel Poverty Action Lab  
Source: 3ie

er words, not all the projects and programmes of an organisation can be evaluated with regard to their impact. For example, statistical approaches allow us to estimate the effects of specific interventions on precisely defined development outcomes, but are of little use when it comes to broad, long-term effects at aggregate levels, such as GDP growth. The time and effort required to track and measure the effect of, say, a microcredit loan on the livelihoods of individual farmers 20 years down the line far outweighs the usefulness of that information – to say nothing of country-level effects.

## WHAT DOES THE FUTURE LOOK LIKE?

The debate about the relevance of impact evaluations fits into a larger discussion about various approaches towards poverty reduction. The economists Esther Duflo and Abhijit Banerjee view the role of development practitioners as analogous to “plumbers” – they argue that incremental fixes to incentive schemes and government service delivery systems add up to substantial improvements in the lives of the poor. From this perspective, impact evaluations are indispensable, as they are ideal for identifying small improvements and can be used to guide small adjustments in programme delivery.

Others argue that a focus on “plumbing” runs into the danger of missing the bigger picture, i.e. the root causes as to why some countries are able to escape poverty while others remain poor. Impact evaluations are poorly equipped to evaluate large, macro-sized programmes, structural change, regime changes, or large reforms such as trade liberalisation.

In sum, rigorous evaluations are always time-intensive and mostly costly. In the narrow context of the programme being evaluated, a rigorous impact evaluation is an imperfect instrument for accountability due to its high cost and long timeline. But impact evaluations do offer indispensable lessons on what works and what doesn't; it goes without saying that they are invaluable knowledge if we are to build more effective development programmes and spend the limited resources we have for poverty reduction more effectively and responsibly. Even if development organisations choose not to conduct impact evaluations themselves, they must know and make use of the ones existing in their field.

**Bartłomiej Kudrzycki** is a PhD student at the Center for Development and Cooperation (NADEL) at the Eidgenössische Technische Hochschule (ETH) Zurich, Switzerland. He is conducting research on youth labour markets in cities, focusing on the context of Benin.

**Isabel Günther** is a Professor of Development Economics and Director of NADEL at ETH Zurich. Her current research focuses on technologies for poverty reduction, analysis and measurement of inequality, urbanisation and population dynamics, sustainable consumption, and aid effectiveness. She has conducted most of her research in countries and cities of sub-Saharan Africa. Contact: isabel.guenther@nadel.ethz.ch

For references and more information on the projects, see the online version of this article at: [www.rural21.com](http://www.rural21.com)



## RCTs AND RURAL DEVELOPMENT – AN ABUNDANCE OF OPPORTUNITIES

No, it is not a Root Canal Treatment, although some sceptics view them similarly. RCT, or Randomised Controlled Trial, is an evaluation design which when applied to the right type of questions, done in the right way, supplemented by alternative methods and interpreted correctly yields probably the most robust results any evaluation design can achieve, and this with less pain (dental or otherwise) than some of the alternatives. An overview on challenges and opportunities.

By Marie Gaarder\* and Sven Harten

Randomised controlled trial (RCT) is one of a range of designs known as impact evaluations whose explicit purpose is an analysis of attribution. The name is actually misleading as it could be understood as designs that exclusively measure impact indicators. Impact evaluations should more appropriately have been titled *counterfactual evaluations* or *attribution analysis*. Be that as it may, attributing an effect (be it an output, outcome or impact) to an intervention means that all other factors outside of the intervention that could also influence the outcome are held constant (or in the jargon “are controlled for”). Typical factors in the rural environment that need to be controlled for are season, weather, access or distance to markets and market prices.

### A ‘WITH VERSUS WITHOUT’ ANALYSIS

So how do impact evaluations, and RCTs more specifically, control

for these factors? Impact evaluations look at what difference a programme or intervention made: did it lead to measurable improvements on some outcome of interest, be it latrine-use in rural India, higher farm incomes through weather index insurance in Tanzania, or increased women empowerment through mobile money (digital financial services) in Northern Uganda? Impact evaluation is a ‘with versus without’ analysis: what happened with the programme (which is a factual record) compared to what would have happened in the absence of the programme (which requires a counterfactual). Most development agencies produce reports about implementation and results at the time of project closure, if not earlier. Why are these usually misleading when it comes to results? These reports typically rely solely on information and monitoring data provided by the programme, and thus

quite frequently fall victim to the before-after fallacy. Consider measuring an outcome both before the programme starts and after it has been implemented for a while.

Typically, if there is an improvement, the programme manager considers the intervention a success. But over the period of any programme, many other factors come into play, not least of which all the other programmes that are being implemented in the same country. Without a valid counterfactual, there is no way of knowing whether the improvement can be attributed to the programme’s activities or may have happened in spite of these. Moreover, spending money on anything or conducting any kind of activity produces some positive effect in many cases: when farmers are trained on a new, semi-automatic irrigation technique, at least some of them will change their behaviour and get some better yields. But was this the most effective and efficient way to increase yield or could a training on improving traditional irrigation have produced much



better results? Then again, if there is no measurable improvement, it could well be that the programme acted as a safety net if for example the same outcome worsened in the rest of the country.

## A CHALLENGING TASK

In designing an impact evaluation, it is important to carefully consider first what is already known (no need to reinvent the wheel), what the important questions are that the programme implementers and wider development community want answered (are they interested in effectiveness? – compare DEval’s Evaluation Programming and Reference Group Model), and how much time and resources are available. RCTs are data-intensive and hence relatively expensive (but not necessarily more so than alternative designs). Designing an under-powered RCT, which has a too little sample size to detect statistically significant effects, is therefore not an effective use of resources. If indeed, as it unfortunately mostly continues to be the case in development programmes, we still do not know whether the type of programme/intervention or some sub-activity of it ‘works’ or has important adverse side-effects, an impact evaluation may be called for. But bear in mind that depending on the effects of interest, these aspects may take time to emerge and to be discernible in the data.

To give an example, agricultural productivity effects will at least take a year to detect (next harvesting season of similar type), while to find out the effect on employability of early childhood development interventions will take one generation (15–20 years). Once we have established that a counterfactual analysis is desired, the next issue to consider is how to establish a counterfactual that best mimics the population that was targeted by the intervention, while taking into consideration what is ethical and feasible in the particular context.

## SPECIFICS OF THE RURAL CONTEXT

There are a number of variants to RCTs that distinguish themselves through the unit of randomisation, the rule applied to assign the population to treatment or control, and the ways in which the treatment is allocated or spaced in time. Each of these will be introduced through an example from the rural development field. It is worth noting that in the databases of the International Initiative for Impact Evaluation (3ie), a large number of RCTs can be found in the rural space.

Based on our experience, there are three issues that distinguish rural RCTs from urban ones. One the one hand, they are easier to implement in the rural context as threats of contamination are relatively low due to limited transmission of information, which is typically contained within the villages. On the other, two issues may be complicating factors in the rural context: first, responses to survey-questions may be more prone to various types of response biases (e.g. social desirability bias), second, depending on the type of intervention, sampling may be more complicated since villages or individual farms are often located in particular micro-climates, soil quality and access to ground water which are hard to detect and measure. In Rwanda, for instance, two apparently identical villages may be in separate valleys within just five kilometres distance, but subject to very different climate conditions. In the rural region around Cochabamba (Bolivia), soil quality at one side of the road may be very different than at the other side.

## RCTs IN PRACTICE

An ongoing 3ie funded study uses individual randomisation to examine the impact of a hybrid risk mitigation financial product that combines credit and insurance, called Risk Contingent Credit (RCC), in rural Kenya. 1,500 households were randomly assigned to receive one of three treatments – (1) the RCC, (2) traditional credit and (3) no credit. The randomisation was done through a public lottery at village level, and the villagers thus knew the treatment status of every participating individual. Individual randomisations are relatively low-cost as sample size requirements are lower. However, such a randomisation within a village faces threats to internal validity. The first risk, namely that of contamination by having individuals ‘switching’ their treatment status, was addressed by making the insurance/credit contracts non-transferable. The other main threat is known as the John Henry effect, when the control group changes its behaviour due to knowledge of what is happening in the treatment group, and in this case, for example, is triggered to seek traditional credit from other banks operating in the region. While the research team may not be able to control this, by collecting information on the credit and source of loan, they will be able to identify and assess the magnitude of this problem.

The second type of example is of cluster-based randomisation, a cluster being a grouping of individuals or households at a level which



A latrine project in Bangladesh. Randomisation has to be adjusted to the level of implementation (e.g. community/village instead of individual level).

Photo: Mirva Tuulia Moilanen/The World Bank

makes sense from a point of view of the intervention and outcome of interest (e.g. village, schools, health centres). Many interventions are implemented for example at the community/village level, with the expected benefits also to be captured at that level, requiring village- instead of individual-based randomisation. The other reason for a clustered approach is the large expected spillovers within tightly knit rural communities, which would entail that other individuals within a community where some members are participating in a programme may also benefit from the intervention (e.g. by watching their neighbours and talking to them). 3ie is funding four ongoing impact evaluations focusing on promoting latrine use among rural households in four different states in India. As all four projects are complex interventions involving social demonstrations, workshops, community events and mixed-media communication, there is a high risk of spillover effects of the interventions among individuals and households in the treatment areas. As a result, all four projects have taken a clustered RCT approach in order to avoid these effects. The Odisha team has randomised at the village level while the Karnataka team is randomising at the Gram Panchayat (the village council, which is the lowest administrative unit in rural areas). The projects have chosen to randomise at different levels given differences in distance.



Where withholding the programme from any group of potential beneficiaries is not desirable, stepped wedge cluster randomised trials can be suitable.

Photo: Jörg Böhling

The previous examples have all been designed to respond to the question ‘does the programme work’ by having control groups that do not receive it. Quite often, however, what you want to test out is modifications to an existing programme to see whether adding a design component leads to improved effectiveness or by comparing different additions to judge which is the more effective innovation. This was the background for a 3ie-funded ongoing multi-arm RCT to test innovative modules of farmer extension services and their effect on agricultural productivity in Cambodia, within the Project for Agricultural Development and Economic Empowerment (PADEE). The authors investigate the impact of two additional features to the traditional extension worker model that provide agricultural advice. First, they assess the impact of incorporating Information and Communication Technologies (ICTs) to overcome extension agents’ low levels of technical education and training. The extension agents are provided with tablets equipped with specialised software with information about soil testing, seed recommendations, fertiliser application, and identification and treatment of crop diseases. Second, the authors test whether performance-based incentives can incentivise extension workers to make use of information available in the software to increase their effectiveness. The authors assess the impact of these features

by randomly assigning a group of 20 villages to each of the following treatment arms: 1) regular extensions services, 2) ICT extension, and 3) ICT plus incentives extension. By measuring the value added of components 2) and 3), they compare the effect of the second treatment to the first and the third treatment to the second and first.

Sometimes, programme implementers are interested in whether the dosage of a treatment makes a difference to the measured outcomes. Stepped wedge cluster randomised trials allow for controlling for variations in timing due to random and sequential crossover of clusters from control to intervention. A 3ie supported impact evaluation in Sudan assesses the impact on incidence and prevalence of moderate acute malnutrition (MAM) in children under five years and pregnant and lactating women of different MAM treatment and prevention interventions. The evaluation design uses variation in the timing of introduction of MAM prevention components (such as food-based prevention, behaviour change communication) and home fortification to localities (clusters) where treatment activities were underway. The impacts are assessed by undertaking a cross sectional comparison across clusters, as well as a comparison over time within the same cluster. This is a good example of a methodology that can be employed for robust causal analysis when baseline data are not available, and where withholding the programme from any group of potential beneficiaries is neither desirable nor feasible.

## ETHICAL CONSIDERATIONS

It is sometimes critically viewed that impact evaluation designs require that only some individuals receive the intervention and this brings up ethical concerns. However, randomisation does not necessarily drive the fact that only some individuals receive an intervention; they are particularly well-suited when for financial or logistical reasons the implementation and roll-out is slow or staggered, or when comparable groups are left out for other reasons. This is the reality of most development interventions. Part of what underlies the ethical concern about impact evaluations is the premise that assignment to a comparison or control group implies ‘not receiving a benefit’. This is not necessarily the case for two reasons. First, the comparison group can be receiving a treatment with which another competing intervention is being compared, as we saw in the multi-arm RCT. Second, it is important to examine the assumption that receiving a

development intervention, or more of one, is always a benefit. The reality is that the effectiveness and impact of a large number of development interventions have yet to be proven. When a genuine state of uncertainty exists about the benefits of an intervention, so that in theory it could be harmful or ineffective, there is an urgent need for it to be critically examined. This state of uncertainty is known as equipoise in the medical literature.

On the other hand, in cases where a programme cannot be implemented across all individuals immediately, randomisation of eligible individuals can in fact be perceived as more ethical and transparent than any other allocation mechanism. While the ethical concerns may sometimes be misplaced or exaggerated for the reasons just described, it is nevertheless critically important to always carefully consider the potential ethical issues that may arise when designing and conducting RCTs.

To summarise, the gaps in knowledge about what works when and where in the rural and agricultural development space (check out the evidence gaps in 3ie’s Evidence Gap Maps) are still immense, and the opportunities to utilise RCT-type impact evaluations to answer effectiveness questions abound.

---

**Marie M. Gaarder** is Director of the Evaluation Office and Global Director for Innovation and Country Engagement at the International Initiative for Impact Evaluation (3ie). She has also held senior positions at the World Bank and Norwegian Agency for Development Cooperation. Marie Gaarder has a PhD in Economics from University College London. Contact: [mgaarder@3ieimpact.org](mailto:mgaarder@3ieimpact.org)

**Sven Harten** heads the Competence Centre for Evaluation Methodology and is Deputy Director of the German Institute for Development Evaluation (DEval). He has more than 15 years of experience as a professional evaluator and most recently worked as Senior Evaluation Specialist at the World Bank (IFC). Sven Harten has a PhD in Political Science from the London School of Economics. Contact: [sven.harten@deval.org](mailto:sven.harten@deval.org)

\* With input from Bidisha Barooah, Neeta Goel, Shaon Lahore, Diana Lopez-Avila, Emmanuel Jimenez, Monica Jain, Tara Kaul, and Francis Rathinam.

For a list of references, links to the studies and further information, see the online version of this article at: [www.rural21.com](http://www.rural21.com)



RCTs are widespread in the medical arena. Here, the environment can be fully controlled.

Photo: WHO/Eduardo Soteras Jalil

## RANDOMISED CONTROLLED TRIALS – THE GOLD STANDARD?

Although randomised controlled trials are seeing widespread use, they have also been in for some criticism. Our author shows some of the snags that the method may meet with and recommends that context and appropriateness be given more consideration in designing evaluations.

By Maren Duvendack

Randomised controlled trials (RCTs) have recently grown in popularity. The basic idea is simple. In a randomised study, individuals are randomly assigned to so-called treatment and control groups, whereby both groups must be drawn from individuals whom the programme has yet to serve, so that the impact of an entire programme can be evaluated.

This random assignment to either treatment or control groups ensures that potential outcomes are not contaminated by self-selection into treatment. Self-selection refers to individuals selecting themselves into participating in particular programmes, e.g. they may self-select into microfinance programmes because they are particularly entrepreneurial or have certain risk attitudes and/or business skills. If randomisation is successful, it is assumed that individuals in treatment and control groups are

equivalent in terms of observable and unobservable characteristics, with the exception of the treatment status. As a result of this, the differences we observe in the outcomes of each of these individuals are understood to be the effect of the programme.

### THE CRUCIAL ASPECT OF CAUSALITY

Hype surrounding RCTs has led policy-makers, funders and researchers to believe that randomisation is the only method that convincingly establishes causality. However, for RCTs to convincingly establish causality, they need to be implemented properly. In other words, we have to be convinced that individuals have been truly randomly allocated to treatment and control groups; only then will we have succeeded in constructing

an accurate counterfactual scenario (i.e. what would have happened in the absence of a programme). At the same time, we must be able to check for self-selection bias without having to resort to sophisticated econometric techniques that require particular technical expertise.

RCTs may be an attractive methodological option but they are not free from challenges, which can be of technical, ethical and/or practical nature. In academic circles, the chorus of critical voices has become louder arguing that there are threats to the internal and external validity of RCTs. For instance, how much can we really trust the causal claims of RCTs, and how generalisable are their results to other situations and/or individuals? Let us now look at some of these threats before examining potential alternatives to RCTs.

## THE CHALLENGES OF CONDUCTING SUCCESSFUL RCTS

Successfully implementing RCTs is not an easy task, mainly due to technical challenges such as ensuring double-blinding, avoiding pseudo-random methods, addressing attrition and considering behavioural changes caused by the experiment itself such as Hawthorne and John Henry effects which may affect the results in positive as well as negative ways (as explained below). Furthermore, spill-over effects cannot be fully ruled out, and ethical and practical challenges need to be considered. We will now investigate some of these challenges in more depth and start with the key feature of RCTs, which is double-blinding.

Evaluation expert Michael Scriven, among others, stresses that double-blinding is one of the prerequisites for a robust RCT. Double-blinding implies that individuals participating in the RCT and researchers executing the RCT do not know who is receiving a particular treatment or not. The rationale for striving to achieve double-blinding is to avoid biased research outcomes caused by the placebo effect. In the medical arena, where RCTs are well established, double-blinding can be ensured by running RCTs in laboratories, where the environment can be fully controlled, but the case is different for studies in the area of the social sciences and international development in particular. For example, RCTs evaluating the impact of education, social services or microfinance programmes are usually not even single-blinded but essentially 'zero-blinded'. In other words, individuals usually discover whether they belong to treatment or control groups, which undermines the notion of double-blindedness.

Another challenge is the prevalence of pseudo-random methods which often occurs during the process of assigning individuals to treatment and control groups. It pays to investigate how exactly individuals were assigned to their respective groups; was the underlying process truly random? For example, the evaluation of the Girl's Education Challenge in Mozambique, funded by the UK's Department for International Development (DFID), claimed to be a RCT but upon further investigation and discussions with the evaluators, it became apparent that some non-random elements had crept into the allocation of individuals to treatment and control groups through challenges encountered during fieldwork. This can obviously have serious consequences for the reliability of the estimates obtained from RCTs, and it is not unusual for studies not to

describe their randomisation process accurately, or in much depth.

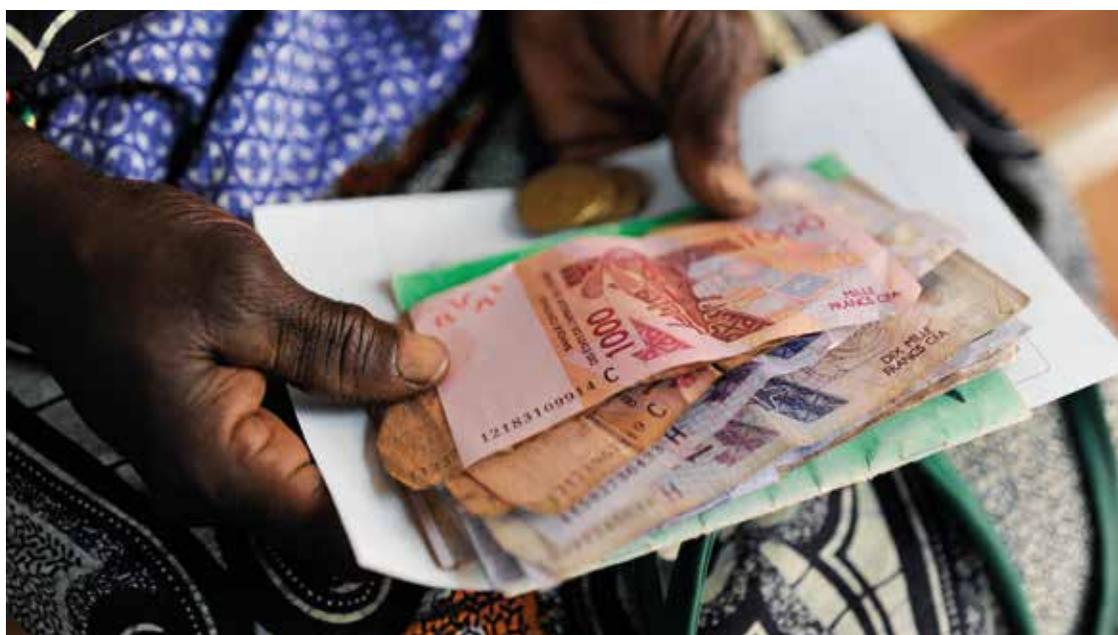
Furthermore, many RCTs do not address the issue of attrition appropriately. Attrition refers to individuals that have been assigned to either treatment or control groups but have then decided not to proceed with the experiment. It is often not clear why those individuals drop out, and this behaviour can have adverse effects on the results of the experiment. It is frequently argued that individuals dropping out would have been worse off than the ones remaining and hence a risk of overstating impact estimates exists, but the opposite can also be true. Drop-outs change the composition of treatment and control groups thereby influencing the results of the experiment since their outcomes cannot be observed. It is possible to track the individuals that drop out, and thereby one can address any side effects of attrition, but this is a costly undertaking. More importantly, all randomised studies should report the level of attrition and compare drop-outs with the individuals that remain in the study to gauge whether there are systematic differences between these two groups – at least in terms of observable characteristics.

Another key challenge affecting the generalisability of RCTs is linked to behavioural changes that can influence treatment and control groups. These behavioural changes are known as *Hawthorne* and *John Henry* effects, with Hawthorne effects referring to behavioural changes in the treatment group while John Henry effects relate to behavioural

changes in the control group. For example, individuals in the treatment group might positively change their behaviour for the duration of the study as they feel thankful for receiving treatment and as a response to being observed. The same behavioural changes might apply to members in the control group altering their behaviour positively or negatively.

A final technical challenge we need to understand is related to spill-over effects that can have adverse effects on the impact estimates obtained from a RCT. Spill-over effects refer to individuals in the control groups that are affected by the treatment in physical ways or in the form of price changes, learning or imitation effects. But individuals in the treatment group can also be affected by spill-overs, e.g. changes in migration patterns through being attracted by the treatment can have an effect on the impact of the programme. In the case of Mexico's PROGRESA conditional cash transfer programme, spill-over effects caused by migration were detected, but the good news is that these spill-overs, if significant, can be measured and checked for. For example, the level of treatment exposure within groups can be adjusted to assess the magnitude of potential spill-over effects.

In addition to these technical challenges, potential ethical challenges should not be overlooked. The implementation of RCTs is not always feasible because of ethical considerations, e.g. how can it be justified that certain individuals are assigned to a treatment group while others are excluded from a potentially



Double-blindedness is usually not possible in evaluations such as those on the impact of microfinance programmes.

Photo: Jörg Bötting

beneficial treatment. Many argue, however, that these ethical concerns are not valid considering that if a treatment is proven to be beneficial, it will eventually become available to all individuals in the control group as well.

Finally, there are practical challenges to overcome in the successful implementation of RCTs; extensive co-operation from the programmes that are being evaluated is required. This can be time and cost intensive. Laura E. Bothwell and co-authors argue that RCTs are high-cost and high-value marketing tools and hence value for money will need to be carefully considered before embarking on one, e.g. with regard to what percentage of the overall programme budget should be allocated for conducting evaluations.

Are the funds sufficient to conduct a high quality RCT? Is the RCT the appropriate methodological option to answer the questions of interest in relation to its costs? Moreover, for RCTs to work, the environment needs to be rigorously controlled, so that any difference in outcomes between the two groups can be adequately attributed to the impact of the programme. Therefore, applying RCTs is in many cases not desirable or feasible and hence, we need to consider robust alternatives.

## LET'S THINK ABOUT ALTERNATIVES

There is an increasing role for qualitative methods in impact evaluation such as process tracing and life histories but also for experimental and behavioural games as well as for social network analysis, longitudinal studies and other modelling approaches. It is beyond the scope of this article to discuss these alternatives in depth, but it should be noted that strictly quantitative approaches such as RCTs can easily be replaced and/or complemented with cost-effective alternatives that often focus on gaining a better understanding of the causal mechanisms that underpin a particular programme with the objective to unpack its 'black box'.

Given the challenges outlined above, is the recent enthusiasm for RCTs sustainable? In principle, RCTs have the best chance to meeting a range of evaluation challenges such as controlling for selection bias, constructing robust counterfactual scenarios, etc. However, Elliot Stern and co-authors argue that in 95 per cent of all cases RCTs are not feasible

or appropriate. Hence, we maintain that we need to think more seriously about alternative as well as complementary methods to RCTs.

RCTs promise rigour and certainty which may explain why they have become so popular but rigour is not just limited to RCTs. Other disciplines such as law, ecology and others rely on other techniques such as rules of evidence, aerial photographs and satellite imagery to demonstrate causation. There may also be value in exploring relatively inexpensive methods that have been little used in the area

**“ We need methodological pluralism and an open-mindedness among researchers and commissioners of evaluation research. ”**

of impact evaluation so far such as experimental and behavioural games, social network analysis, agent-based modelling and other simulation approaches. These approaches can often be more powerful than RCTs alone for understanding the underlying causal mechanisms of programmes, and they are particularly useful when faced with small n evaluations (those involving small sample sizes) and/or evaluations of complex interventions in particular in conflict-affected areas where RCTs have serious limitations.

The choice of an evaluation study design, whether to use a RCT, a quasi-experiment, qualitative tools or a mixture thereof, should depend on the objectives of the evaluation, access to financial resources and time horizons. Methodological rigidity will not help us to better understand the effectiveness of development programmes – what we need is methodological pluralism and an open-mindedness among researchers and commissioners of evaluation research to allow the best possible evaluation design given the specific context we find ourselves in. Context and appropriateness of methods matters!

**Maren Duvendack** has a PhD in development economics from the University of East Anglia (UEA), UK. Her key research areas cover applied micro-econometrics, impact evaluation methods, systematic reviews and meta-analysis, microfinance, replication and reproduction of quantitative analyses as well as research ethics. Duvendack has extensively worked on microfinance impact evaluations in India and Bangladesh. Contact: m.duvendack@uea.ac.uk

For a list of references, see the online version of this article at: [www.rural21.com](http://www.rural21.com)

## WHAT ABOUT EVALUABILITY?

All considerations regarding the right design of an evaluation aside, one aspect that must not be forgotten is evaluability, i.e. “the extent to which an activity or project can be evaluated in a reliable and credible fashion”, as defined by the Development Assistance Committee of the Organisation for Economic Co-operation and Development (OECD-DAC).

While an evaluation aims to judge the merits of a particular intervention, an evaluability assessment occurs before an evaluation. It can support formulating a recommendation on whether an evaluation is worthwhile in terms of its likely benefits, consequences and costs. Also, it can show at which point the evaluation should take place and help decide whether a programme or intervention needs to be modified, whether it should go ahead, or whether it should be stopped. Assessing the evaluability of a measure can prevent wasting valuable time and resources on a premature or inappropriately designed evaluation. And, as a WorldBank Group blog explains, it can “thwart ‘evaluitis’ and the ‘ritualization’ of evaluation processes”.

The Overseas Development Institute (ODI – UK) authors of the manual “Evaluability Assessment for Impact Evaluation” maintain that three focus areas ought to be covered by an evaluability assessment:

- the adequacy of the intervention design for what it is trying to achieve,
- the conduciveness of the institutional context to support an appropriate evaluation, and
- the availability and quality of information to be used in the evaluation.

The guide contains a checklist to help evaluators to answer the following key questions:

- 1. Is it plausible to expect impact?** This is where the adequacy of the intervention design is examined. Do stakeholders share an understanding of how the intervention operates? Are there logical links between activities and intended impact?
- 2. Would an impact evaluation be useful and used?** Here, the focus is on stakeholders, demand and purposes. Are there specific needs that the impact assessment will satisfy, and can it be designed to meet needs and expectations?
- 3. Is it feasible to assess or measure impact?** This question refers to data availability and quality. Is it possible to measure the intended impact, given on-the-ground realities and evaluation resources available?

The manual is available for downloading on the ODI website: [www.odi.org](http://www.odi.org). Useful information on evaluability can also be found on the BetterEvaluation project website: [www.betterevaluation.org](http://www.betterevaluation.org) (sri)

## LEARNING FROM PARTICIPATORY EVALUATIONS

Evaluations have many aspirations, not only to account for results. Evaluations are tools for reflection and learning, for decision-making and team-building, for ownership creation and empowerment, and for advocacy. The current – and no doubt indispensable – debate on providing robust evidence on the impact of development interventions tends to lose sight of the valuable multifunctionality of evaluations.

By Kai Schrader

The discussions around rigorous evidence on the impact of development programmes and the use of accurate scientific methods tend to veil the subjectivity of impact evaluations; the moment, the research subject, the methods, the “deliverables” and the participants, both evaluators and evaluated, are the result of intentional multistakeholder decision-making. There is a purpose behind each evaluation, which needs to be outlined in the description of the process and methodology applied of an evaluation. Since impact evaluations are costly investments, donors certainly play a crucial role regarding the type and quality of such exercises.

Helvetas, as a “Learning Organization”, endeavours to strengthen the learning aspect in most of its undertakings, methods and tools, evaluations included. Convinced that learning is, as described by Harold Jarche, an advisor on organisation development, “a continuous process of seeking, sensing, and sharing” and happens through participation, engagement and communication, we support our partners and staff in leveraging their rich and diverse knowledge by fostering critical reflection and exchange. In the field of project or programme evaluations, we therefore explore, promote and apply methods that bring in the knowledge and perspectives of stakeholders at various levels. Some of these are described in the following.

### PRIMARY STAKEHOLDER AND LOCAL INSTITUTIONS – SOCIAL AUDITS AND “BENEFICIARY ASSESSMENT”

Social Audit, the assessment of the performance of “duty bearers” – e.g. public services of local governments – carried out by the “right holders”, citizens or users of such services, is an evaluation method that improves “downward” accountability and, finally, the quality of public services. Social Audits as well as Client Satisfaction Surveys are useful learning tools in projects that support local public institutions in discharging their responsibilities in delivering quality public services and



Community representatives in Nepal (Thakur Thapa) answering people’s questions during a Public Audit.

Photo: HELVETAS Swiss Intercooperation

respond to citizens’ needs. The strengthening of such participatory and inclusive evaluation practices contributes to creating processes for dialogue between stakeholders that per se are results of development, as we have observed in Eastern Europe and Bangladesh.

Another approach or methodology which also fosters the empowerment of primary stakeholders is the “Beneficiary Assessment”, as it has been known since its description by then World Bank’s Lawrence Salmen in the 1990ies. It is a qualitative method used to improve the impact of development operations by capturing the views of intended beneficiaries regarding a planned or ongoing intervention. Community members, farmers or other project participants are trained as peer observers in a two- to three-day workshop. They then identify the research questions and enter a process of interviewing peers in their communities. The objective of this method is to assess the value of an activity as perceived by project beneficiaries and to integrate findings into project steering. It is designed specifi-

cally to undertake systematic listening of the project participants and other stakeholders by giving voice to their priorities and concerns. This method of systematic consultation is used by project management as a design, monitoring, and evaluation tool. The WARM (Water Resources Management) project in Nepal (see Box on page 17) shows that the nature of Beneficiary Assessment provides the development of peer observer skills, which are appreciated by the participants and communities and can benefit other development processes such as identification and planning of new interventions.

### OWN STAFF AND PARTNERS – PEER REVIEWS AND “CAPITALISATIONS” FOR KNOWLEDGE SHARING

External assessments of interventions are valuable and indispensable pieces in Project Cycle Management. While final evaluations are usually performed by external collaborators, project Mid-Term Reviews can be conduct-



Stakeholders evaluating a Farm Water Management project in Kyrgyzstan.

Photo: Stefan Stolle



## RESULTS OF PARTICIPATORY IMPACT ASSESSMENTS (PIA) IN WATER AND SANITATION PROJECTS\*

### “Water Resource Management Programme in Nepal” (WARM-P), HELVETAS

- Clients of water supply services report improvement of water quality and time savings.
- 80 per cent of households have a toilet and use it.
- Focus group participants indicate no discrimination based on caste and economic hierarchy.
- Need to strengthen the behaviour change of hygiene and sanitation.
- Stakeholders appreciate project’s overall participatory approach of identifying WASH (Water, Sanitation and Hygiene) priorities.

\* Studies conducted in 2013

### “Rehabilitation and Improvement of Water Sources in Borana, Ethiopia”, HEKS & OSHO

- Community members played an active role in identifying water sites for rehabilitation and during the construction process.
- Water is considered a shared property for everyone, irrespective of social status.
- Perceived changes were shortened distance to water source and year-round availability of water.
- Improved human health by better water quality, but no difference in animal health; participants requested additional water schemes.
- The process was considered empowering, and participants of validation workshop unanimously called for the use of PIA in the future.

ed without external support by own thematic advisors and involve staff and partner organisations. Although this is a lighter and more flexible process, it often brings valuable findings as participants tend to be more engaged and more critical of their own performance than external evaluators who might need to communicate more cautiously. Moreover, the teams joining a self-assessment exercise gain ownership on the findings, conclusions and recommendations. Their contributions to the procedure translate into stronger commitment towards complying with agreed follow-up actions.

We also endorse processes that bring in stakeholders, professionals, partners and colleagues at different stages for evaluations of highly complex operations or higher-level evaluations, for example for sector- or country programme evaluations. Peer reviews focus on the facilitation of such endeavours and bring the people together, be it face-to-face and synchronous or at distance and nonsynchronous. The multiple perspectives and expertise of involved individuals and the numerous insights and opinions enrich such evaluations and foment learning among peers; this is a moment of intense knowledge sharing. Bringing together and showcasing the experiences of all colleagues participating in bio-cotton-projects (see Rural 21 no 2/2017) or interventions that focus on building up Rural Advisory Services offers learning opportunities for one’s own and many other organisations.

The Review of Country Strategies can bring in the valuable knowledge of many colleagues when facilitated in a participatory way. Country directors and thematic experts from neighbouring countries can contribute to analysing the progress of a country programme, and learning takes place at various levels and on

‘both sides’, evaluators and evaluated. This benefits mutual inter-organisational learning, the sense of belonging to one organisation as well as the regional focus of collaboration, as we have seen for example in Central Asia.

Another promising evaluation procedure after finalising longer interventions is the Capitalisation of Experience (CAPEX), which allows to gather and systematise all relevant project documents and collecting of insights from participants and external key informants on good practices, failures and lessons learned, as we recently did for our engagement in Bhutan’s Community Forestry Sector. CAPEX publications are important sources of information for strategic decision-making as well as for interested persons in the sector and the region alike. They are shared on relevant platforms and networks and are useful “certificates” for staff to show their career and expertise to others.

### LOOKING AHEAD

Internationally, there is a trend to professionalise evaluations and to put the decision on design, methods, indicators and measurement into hands of the academia. The interaction of NGOs with research institutions to understand impact is a very welcomed and fertile evolution. Scientific research in development projects is a complementary procedure of participatory methods, and not a substitute. Given its high costs and tardy results, rigorous assessments are exceptional studies that should be conducted regularly and well-planned in selected projects. But development organisations need information on impact of all their projects, at early stages of implementation and in a useful format for decision-makers in the countries.

Helvetas’ M&E strategy for improving result orientation and impact is to build up capacities for better evaluation in the teams that perform regular project M&E through training and coaching; to lift their attention from activities and expenditures to outcome and impact through improving reporting; to improve indicator selection and measurement methods at the very beginning of a project; and to make M&E slimmer and more useful. Finally, development organisations are also accountable for spending on M&E and impact assessments, which need to be justified by their usefulness for learning and steering, improving future actions and developing capacities in the countries. We therefore strive to mix methods and evaluation designs, to adapt the evaluation process and participants to the context and specific situation of the project and to be flexible and innovative with applied methodologies. The critical analysis of the “evaluability” of a project, the participatory process of defining methodology and timing, and the involvement of staff and partner organisations in impact evaluations contribute to capacity development, empowerment and learning.

**Kai Schrader** is advisor for Evaluation & Learning at HELVETAS Swiss Intercooperation. Having worked in development co-operation for over 20 years, he has focused his activities on participatory methods for planning and evaluation, rural development, and topics related to land use, agriculture, and ecology. Kai Schrader holds a PhD from the Centre for Development and Environment of the University of Berne.  
Contact: kai.schrader@helvetas.org

For more information on the Helvetas projects, see: [www.rural21.com](http://www.rural21.com)

## MAKING PEOPLE VISIBLE

National offices of statistics as well as monitoring systems of development agencies tend to inform at household, farmer or family level, but often don't go any further. To allocate resources, specifically to women, men, girls and boys, in a meaningful way, it is essential to make people visible by using gender and age group sensitive language and by describing the conditions and context under which these different groups of people are living.

By Felix Fellmann, Sophie Hirsig and Ueli Mauderli

Poverty reduction is at the core of sustainable development. Almost half of the world population – more than three billion people – still live in poverty. In order to make progress towards global justice, security and sustainability, decent income and employment for 1.3 billion extreme poor need to be created, and 800 million people suffering from hunger need to reach food security. An analogue challenge was recognised already in 2000, when the Millennium Development Goals (MDG) process of 2000–2015 defined the eradication of extreme poverty as its first goal. Efforts to reach MDG 1 are relatively successful, because of robust results from countries like China, Vietnam, and Brazil. Finally, in September 2015, with the Sustainable Development Agenda 2030, the United Nations agreed to meet SDG 1 “to end poverty in all its forms” and SDG 2 “to end hunger (and all forms of malnutrition)” by 2030.

Multilateral and bilateral development agencies as well as many civil society organisations continue to focus on improving the economic and social situation of women, men and children living in poverty. But what do we really know about these people? To what extent do development agencies know the living conditions of their target groups? And what is the quality of solid gender and age sensitive data that development agencies generate? It is difficult to gain access to such knowledge-effective resource allocation for poverty reduction.

### WHY VISIBILITY MATTERS

Information generated by National Statistics Offices often focus on household (HH) and estimate the average number of people living in a household or their overall household income. Such generic information ignores the most salient issues such as:

- The political economy: To what socio-economic strata the household belongs to?
- Social data: Who exactly is living in a household?

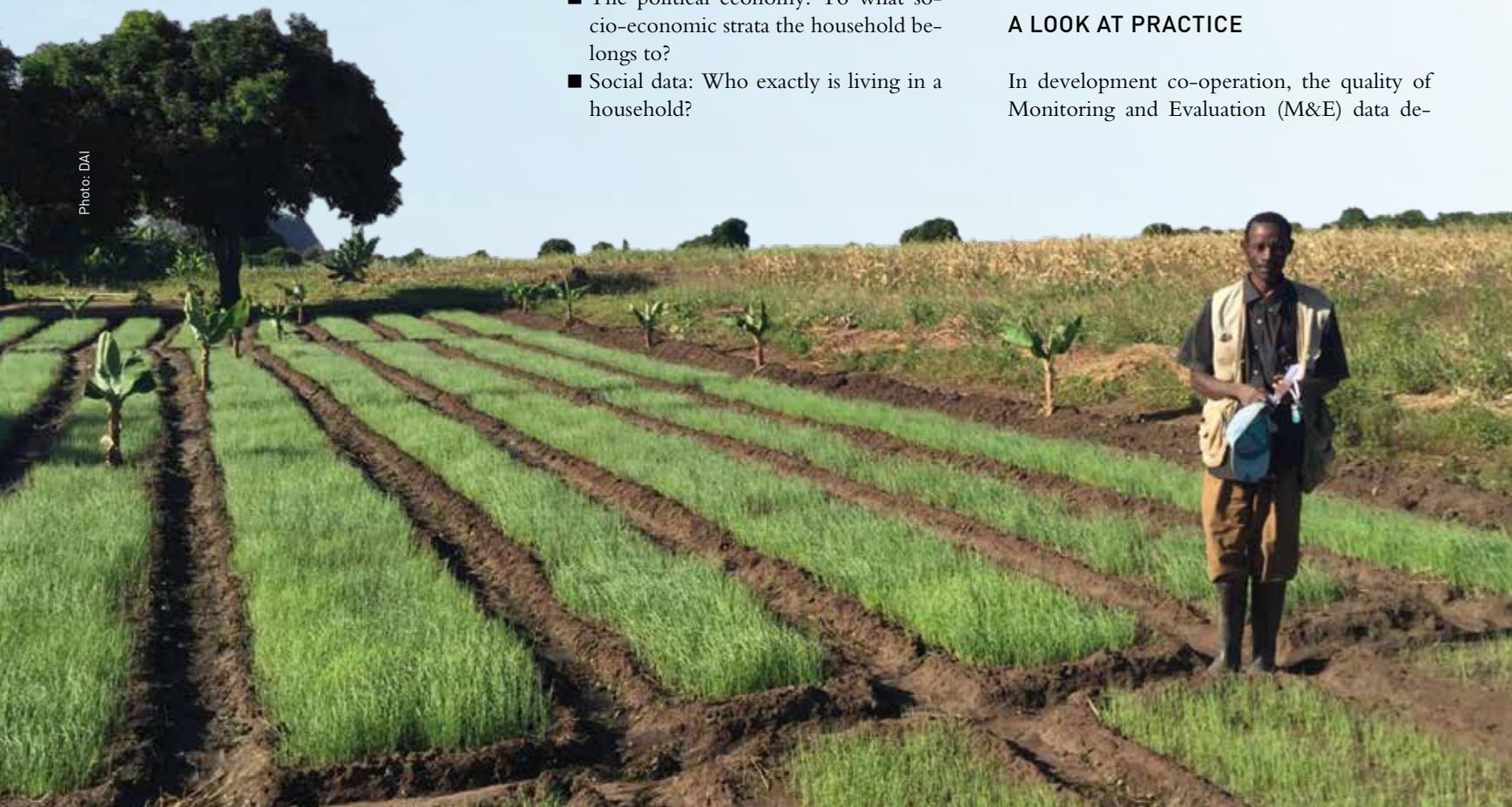
- HH economy: How is the income stream distributed during the year?
- Power relations and gender equity within the HH: How (process) and by whom (men, women) is the household income generated, allocated and used?

Solid data (statistics, large-scale studies) on these questions is difficult to find in spite of the fact that above questions are at the core of human social and economic development. Often, data is scattered, comprises small samples of men and women and lacks triangulation.

Aggregation of such sensitive and vital data is hardly possible, which is a reason why development policy discussions are often insufficiently informed. However if the real people hidden behind our beneficiary numbers or households remain invisible, developmental change will remain slow, and women and men, boys and girls will be left behind.

### A LOOK AT PRACTICE

In development co-operation, the quality of Monitoring and Evaluation (M&E) data de-



defines the degree to which people have been made visible as women, men, boys and girls living in specific social, economic and ecologic contexts or the degree to which those same people are hidden in unspecified households. But what does it look like in practice?

Precisely in order to examine this, the Swiss Agency for Development and Cooperation (SDC) conducted applied research on M&E in Country and Regional Programmes in Africa, Latin America and Asia. Language and formulation of outcomes and indicators were analysed and compared with the annual result reporting. One aim of the study was to identify areas for improvement in order to be able to produce reports with strong content for communication to the Swiss public and to the Government. A second aim was to generate lessons and best practices from the monitoring process, fostering in-house learning. But whereas M&E covers many aspects, the study concentrated on the degree to which people and target groups were made visible and distinct.

21 Country- and Regional Strategies with a strong focus on agriculture and food security as well as the corresponding annual reporting were analysed for the period 2013 to 2016. The Table below shows the areas of observation. It demonstrates that only in a third of all cases were the outcomes formulated in gender-sensitive language. Just a fifth of each indicator was formulated in an explicitly gender-sensitive manner or with an explicit poverty/inclusion focus.

### WHAT DOES INVISIBILITY, RESPECTIVELY VISIBILITY MEAN IN PRACTICE?

Statistics at national level often estimate indicators at household level. Words such as “person”, “family”, “herder”, “farmer”, “entrepreneur”, etc. are used. These terms are also often observed in reports of development projects. However, they are too unspecific to really make qualified assessments regarding the outcomes for the beneficiaries. The Table on page 20 shows the terms with which different degrees of visibility can be reached.

### CHALLENGES FOR M&E

In multiple discussions with practitioners in the context of the studies, a number of challenges have been identified for M&E that are closely linked to the above problem. The most frequent ones concerned the following areas:



Gender and age group-sensitive language is essential to make people visible in M&E.

Photo: Joao Costa/Zwela/InovAgro Mozambique

**Complexity:** Development processes are indeed complex, and this is even more the case in contexts that foreigners are not familiar with. It is demanding and requires the courage to deal with uncertainties and to reduce complexity to feasible and practical interventions. In principle, there is no alternative to reducing complexity and to tackling these issues, which offer highest development benefits (leverage). For development co-operation actors – foreign to the context of intervention – ensuring broad participation of different groups of a local population is the best way to integrate formerly invisible social dynamics and has a higher chance to generate sustainable improvements.

**Result chains (outcome level):** The most intense discussion in intercultural teams arises

in debates over outcomes and indicators. To make people visible, gender differentiation and context qualifiers need to be integrated in the entire result chain, i.e. in the:

1. theory of change,
2. impact hypothesis,
3. outcome,
4. indicators.

The extensive study of the 21 SDC Country/Regional Strategies revealed a high potential and a high willingness of teams to improve quickly in the area of monitoring and reporting.

**Monitoring systems:** Project interventions are frequently planned without adequate monitoring systems to observe change. To make

#### Areas of observation and statistical findings

Number of country/regional programmes in agriculture and food security	21
Number of strategies analysed	32
Number of annual reports analysed	84
Number of outcomes for agriculture & food security investments defined	84
Proportion of outcomes explicitly formulated in a gender-sensitive language	33 %
Proportion of outcomes dealing only with one indicator	40 %
Number of indicators serving the 84 outcomes	284
Proportion of indicators with explicit gender-sensitive formulation	21 %
Proportion of indicators with explicit poverty/inclusion focus	20 %

sure solid observations are being made, monitoring systems need to be relatively simple and require to be tested before being put into operation.

The integration of visibility language and qualifiers (see Table) was tested in a number of programmes and is functioning. Observations have shown that agreed parameters such as indicators or baselines are either used for a longer period of time (time series), are used not at all or only partly. The latter is rather common. Hence, we are confronted with the paradox of wanting to capture complexity on the one hand and of being exposed to the practical limitations of monitoring practices on the other hand.

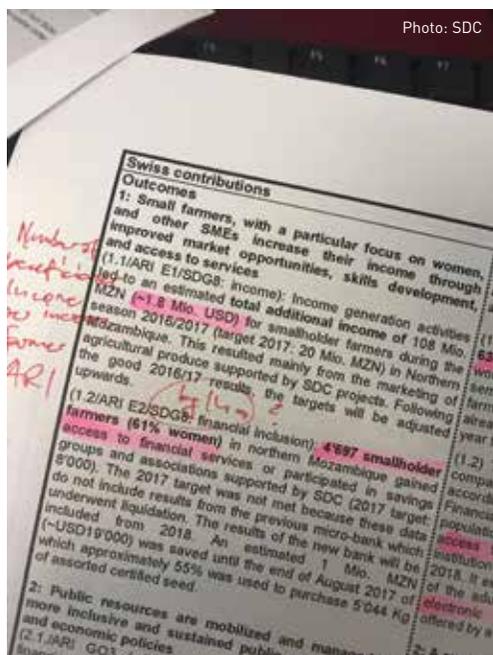


Photo: SDC

**The baseline:** Change is difficult to observe without a baseline (counterfactual). Quantitative changes in particular need points of reference or control groups. However, development projects are designed with a development ambition and rarely follow rigorous scientific designs. On the other hand, there is a legitimate expectation of taxpayers and decision-makers to receive high-quality progress reports. In the absence of baselines and control groups, qualitative sample interviews or focus group discussions – if well done – are producing good evidence on the results of activities as well. But in order to assess stringently the cause and effect of the intervention, a scientific design with control groups is needed. To observe a trend – often required in development co-operation – baselines fulfil the purpose. However, in situations with potential strong externalities like income and crop yields, baselines have strong limitations. Focus

**From invisibility to visibility**

a) Invisibility language	b) Visibility language	c) Context qualifiers
Household	Women	Age
Farmer	Men	Handicapped
Family	Boy	Living in poverty
Herder	Girl	Living with hunger
Entrepreneur	Elderly	Unemployed

The highest visibility is achieved if gender-sensitive language (b) is combined with strong context qualifiers (c). Given the mandate to reduce poverty and improve governance, a high gender and context visibility should be standard in development co-operation and M&E.

group discussions with men, women, girls and boys on aspirations and changes complement the picture and add to a high quality of result observations and reporting.

Finally, to make people visible in the four challenges mentioned above, gender-sensitive language as well as context qualifiers need to be integrated into all four areas – development interventions, result chains, monitoring systems and counterfactual/baseline.

Agency for Development and Cooperation (SDC) and partners.

Contact: Felix.Fellmann@eda.admin.ch

**Sophie Hirsig** works as a Programme Coordinator in the Asia Europe Division in the Swiss Red Cross. She has a MSc in Geography and is committed to improving equality and poverty data as a key factor in achieving impact in all programmes of development co-operation and humanitarian work. She co-facilitated the SDC Agriculture and Food Security Network until July 2017.

Contact: Sophie.Hirsig@redcross.ch

**Ueli Mauderli** has been the new SDC Focal Point of the Agriculture and Food Security network since August 2017. As a Natural Resources Management expert returning from his last assignment in Tanzania, he has witnessed the importance of benefiting the poorest in development activities, if goals balancing economic, social and ecologic sustainability are to be achieved.

Contact: ueli.mauderli@eda.admin.ch

**Felix Fellmann** is in charge of the SDC climate programme in the Embassy of Switzerland in Beijing, China. He is an agronomist by background, has several decades of co-operation experience in African, Asian and Latin American countries and is passionate about improving evidence on the impact of development work. Until 2017, he led the Agriculture and Food Security Network of the Swiss

**FOUR PRACTICAL STEPS TO MAKE PEOPLE VISIBLE**

- 1. Give leadership to those men and women who benefit from the intervention.** Designing and implementing a development intervention that fits the context of people living in this given environment and gives them a role is the most empowering investment with a high visibility.
- 2. Apply the four elements:**
  1. Theory of Change for a chosen topic,
  2. impact hypothesis,
  3. outcomes,
  4. indicators; counterfactuals/baselines for quantitative indicators.
- 3. Apply the “magic words”:** “visibility language” and “context qualifiers”. This is a must in any project design and monitoring system (see Table on top of page).
- 4. Involve women and men in monitoring and make it an empowering and learning process.** Build a monitoring system that fits the context, uses the four elements as well as the “magic words” and links to the realities of men and women wanting to improve their life situation.

# IMPACT ASSESSMENT IN COMPLEX EVALUATIONS

Photo: Cornelis Gollhardt/laif

With the aspect of impact gaining importance in development co-operation, there is a growing demand for evaluations. This article's authors maintain that a more systematic integration of evaluation methods is needed to reflect the complexity of the real world and to make results more meaningful.

By Martin Noltze, Gerald Leppert & Sven Harten

In times of scarce resources and mounting public interest in questions around global development, there is a growing demand for impact evaluations as a means of measuring whether public resources are spent effectively and efficiently. Policy-makers, development partners and implementing agencies want and need to show that they make decisions based on evidence and that they learn from what works and what does not. Stakeholders, including funders, beneficiaries and the general public, increasingly ask for information whether spending was meaningful and effective. Besides this issue of aid effectiveness, accountability and transparency are central to development co-operation. Comprehensive global development agendas, such as the 2030 Agenda for Sustainable Development, emphasise the role of impact evaluations in assessing the achievement of highly aggregated development targets. The past decade has also seen the advent of new actors in international co-operation, such as philanthropic organisations, private sector companies or new forms of social investment funds. What these actors have in common is a firm belief in measuring success (i.e. return on investment) through quantifiable indica-

tors. These developments have multiplied the demand for rigorous impact evaluation, meta-analysis and evidence mapping.

## A CALL FOR A COMPREHENSIVE APPROACH

Yet, despite a long history of interdisciplinary work and many interesting recent developments, the evaluation profession has not sufficiently lived up to the challenge of presenting a comprehensive approach. We therefore call for a more systematic integration of methods with a view to bringing together quantitative measures of results achieved and a thorough understanding of the underlying causal mechanisms. In other words, evaluations must get out of the trenches where either answering the question of “how much” was achieved or “why” and “how” is the predominant focus. Indeed, by building on encouraging theoretical developments and making better use of new types of data, it will be possible to answer both questions in a rigorous fashion. In order to make this argument, we will first recapitulate the main contemporary challenges of

evaluating development co-operation and the response of the profession to them. Next we will point briefly to the opportunities presented by “big data” and then make the case for an integrated, comprehensive approach.

With impact evaluations moving more into the spotlight of development co-operation, a range of new challenges are emerging. Firstly, evaluations can no longer hide in a niche of either measuring impacts of individual projects very rigorously or assessing broad programme implementation at a higher level. Indeed, evaluations are expected to go both deep and wide. Secondly, development programmes are becoming more and more complex. Typical interventions include a variety of instruments to reach multi-dimensional development targets implemented by multiple stakeholders. On the one hand, global agendas such as the Paris Agreement and the 2030 Agenda for Sustainable Development defined a large number of detailed impact indicators at a highly aggregated level (see also the article on pages 26–28). On the other hand, both agendas also raise the demand for disaggregated impact statements since the



Testing of a programmed questionnaire for a quantitative household survey in the Philippines.

Photo: Malte Lech/DEval



Validation of information from satellite

“leave-no-one-behind” principle advocates measuring effects at an individual or household level. Thirdly, impact evaluations are supposed to deliver results in a timely manner and thus enhance policy relevance. Whilst all impact evaluations aim to (quantitatively) answer the question “to what extent” results were achieved, the focus has broadened over the past few years to also include questions on “how”, “why” and “under which circumstances” an intervention caused an effect. Evidence-based policy making requires both, knowing the impact and understanding the underlying causal mechanisms.

### REFLECTING THE COMPLEXITY OF THE REAL WORLD

Hence evaluations are facing the challenge of higher expectations regarding the number and types of questions that have to be answered while also backing the answers with quantifiable evidence. In response to the difficulty to meet all these requirements, the evaluation profession deepened the trenches between ostensibly opposing methodological (and epistemological) camps. Part of the profession concentrated on a narrow set of well measurable (i.e. “to what extent”) questions. Others chose a wider set of (i.e. “why”/“how”) evaluation questions in an attempt to better reflect the complexity of the real world. In this view, interventions are part of a causal package and only work in combination with other factors such as the cultural background of the ben-

eficiaries, stakeholder behaviour, institutional settings, environmental context, etc. Building on this, over the last decade, evaluators have constantly been working on broadening the range of impact evaluation methods. Besides randomised controlled trials (RCTs) – the former “gold standard” of the development evaluation community (see also the article on pages 9–11) – other econometric, theory-based, case-based and participatory approaches have gained ground and experienced enormous improvements in the field of systematic testing procedures, a prerequisite for rigorous causal inference.

Beyond the increased variety of rigorous methods, the field of impact evaluation benefits from new forms of data collection and analysis which emerged in the digital revolution era. Both monitoring and evaluation build increasingly on information gathered by mobile technologies, social media and remote-sensing data (see also the article on pages 24–25). On the side of data analysis, so called “machine learning” is especially innovative. Setting out from computer algorithms, machine learning predicts trends based upon the processing of large data sets. On their own, new and larger data sets are not a panacea. Often, they only reflect major trends and probabilities without useful contributions to the questions of attribution and causality. However, great potential lies in the integration of big data and machine learning in complex evaluations and in the triangulation of such data sets with case studies and cross-case analysis.

### MORE THAN COMBINING EXISTING METHODS

In light of the broad range of impact evaluation methods, their strengths and weaknesses, as well as new forms of data, there is a huge untapped potential of integrating multiple methods in complex evaluations. The basic idea of mixed- or multi-method approaches is to overcome weaknesses of one method with the strength of another. However, thus far most of the literature has concentrated on how to combine (or “mix”) quantitative and qualitative methods. In our view, this is too narrow and has resulted in RCTs being complemented by a few focus group discussions or the implementation of a survey as part of a qualitative case study. While this is certainly meaningful for the individual studies, the overarching aim is to achieve more comprehensive evaluation designs that improve the measurement of causal inference and answer complex and multi-dimensional evaluation questions. In this sense, the authors regard comprehensive multi-method approaches as going beyond combining qualitative and quantitative data. Rather, they combine theorising about how activities lead to outcome and impacts (as in theories of change) with different types of causal inference.

For instance, RCTs or quasi-experiments rely on the counterfactual logic, comparing the outcome of one or more treatment groups to the outcome of controls, in other words comparing the beneficiaries of an intervention with



data by aerial photos from a drone in the Philippines.

Photo: Courtesy of Julian Barth/Kigali-Films



Focus group discussions as part of a realist mixed-method design in Ghana.

Photo: Marcus Kaplan/DEval

those not having received the intervention. Statistical models such as longitudinal studies or most econometric techniques draw causal inference from the statistical relationship between cause and effect or between variables. Theory-based approaches include causal process designs that build on the identification of causal processes or causal pathways (e.g. process tracing or contribution analysis) and causal mechanism designs that consider supporting factors and causal mechanisms (such as the realist evaluation paradigm or congruence analysis). Case-based approaches include grounded theory or ethnographic approaches and rely mainly on within-case analysis. Cross-case analysis of several case studies can be managed by configurational approaches such as qualitative comparative analysis (QCA) with the analytic generalisation based on theory. Thereby, integrated approaches are able to combine exploratory and explanatory approaches in sequence or in parallel, nested or balanced with different conceptual frameworks to causal inference. Thanks to their integrative character mixed- or multi-method approaches are also open for new data types and analytical approaches such as geographic data, big data or text mining.

This article argues that no methodological approach is best or even sufficient on its own. Complex development challenges require complex interventions and consequently more complex evaluation designs. Multi-dimensional questions and the need to not only measure the impacts, but also to understand the under-

lying causal mechanisms, require an extension of the toolbox of researchers and evaluators. Facing these challenges for impact evaluation in the field of development co-operation today, only the systematic integration of different forms of causal inference can sufficiently address this demand. Certainly, (quasi-) experimental designs that allow for high levels of rigor and attribution are one important piece in the evaluators' toolbox in complex impact evaluations. However, they are best understood as one of the elements in a complex evaluation design.

The design of advanced mixed-method approaches explicitly follows the particular epistemological interest of the evaluation questions. Through the smart and systematic integration of methods, they are superior to single-method approaches, as they can better address the complexity of interventions, making any discussion on "gold standards" obsolete.

Thus, future development of impact evaluation designs should focus on enhancing the capacities of multi/mixed-method approaches beyond simply sequential or triangulation strategies. At the same time, evaluators should not hesitate to improve systematic testing procedures of single methods to improve the robustness of the overall design. New forms of data collection and analysis raise the bar for possible mixed-method approaches and thus significantly contribute to the further development of future impact evaluation designs.

**Martin Noltze** is a senior evaluator at the Competence Centre for Evaluation Methodology of the German Institute for Development Evaluation (DEval). He holds a PhD in agricultural economics from Georg August University of Goettingen, Germany. Together with Gerald Leppert and Sven Harten, he is co-ordinating DEval's methodology project on Integrating Multiple Methods in Complex Evaluations.

Contact: martin.noltze@deval.org

**Gerald Leppert** is a senior evaluator at the Competence Centre for Evaluation Methodology of DEval, holding a PhD in economics and social sciences from the University of Cologne, Germany. Contact: gerald.leppert@deval.org

**Sven Harten** heads the Competence Centre for Evaluation Methodology and is Deputy Director of DEval. He holds a PhD in political science from the London School of Economics.

From 2018–2021, the German Institute for Development Evaluation (DEval) is implementing a four-year research programme on the integration of multiple methods in complex evaluations. The methodology project will be accompanied by empirical testing of new forms of method integration in several DEval evaluations.

For a list of references, see the online version of this article at: [www.rural21.com](http://www.rural21.com)

## MORE THAN PLUG-AND-PLAY – DIGITAL SOLUTIONS FOR BETTER MONITORING & EVALUATION

A good M&E system empowers projects to appreciate their achievements, learn from challenges and improve strategic decision-making. Yet in practice, project staff often see M&E as an annoying ‘ticking-the-box’ exercise just to satisfy donors. Our author argues that digital tools can make M&E more efficient and even fun if applied with the user in mind. He shares practical insights on how projects can digitise M&E to be both informative and motivating.



Digital readiness and skills are decisive in the choice of a data gathering system. For example, where people rely on simple feature phones, SMS surveys are suitable.

Photo: Jörg Böhling

By Daniel Brumund

It is an open secret that many project managers and staff consider M&E a laborious extra burden that keeps them from their actual work – even if they are well aware of its benefits. Many, if not most, of them would agree that monitoring and evaluating their activities’ progress and results is important for informed project steering, accountable reporting and even for convincing public relations. They might also agree that reliably managing and communicating data helps promote co-operation, co-ordination and transparency among projects, partners, donors and beneficiaries. But if the benefits of M&E are known, then where does the apparent reluctance to tap into them come from?

The experience of Mainlevel Consulting in supporting projects to improve their M&E through digital solutions shows that the reasons often come down to the procedures and mechanisms that projects use to operationalise their M&E systems. A common challenge is the lack of reliable data sources or user-friendly tools to collect, transfer and analyse data. This makes data management and reporting a time-consuming and frustrating affair for anyone involved. Fortunately, this is also where digital solutions bear the most potential to support and replace M&E frustrations with motivation.

### REAPING DIGITAL BENEFITS FOR M&E

Recent years have seen a rapid spread of digital technologies. More and more people, even in

remote areas, own mobile phones giving them unprecedented opportunities to access and share information. This is changing the way sustainable development is informed, mediated and advanced. According to the 2016 World Development Report, digital technologies can make development more inclusive, more efficient and more innovative by facilitating access to information and services, by reducing the cost of economic and social transactions, or by simplifying the upscaling of promising solutions. Projects can greatly benefit from these digital opportunities to improve core M&E processes such as gathering, disseminating and analysing data.

For instance, mobile applications can simplify the collection of monitoring data, especially in decentralised projects. For a civic education project in Lesotho, Mainlevel developed a mobile app enabling staff in district centres across the country to collect data on activities via flexible questionnaires (see screenshots on page 25). This replaces Word files that were sent via e-mail and manually transferred to Excel. Anyone more comfortable with speaking than writing can use voice-recognition to record texts. In remote areas where network coverage is unstable, staff can rely on the app because it is low on bandwidth and securely stores data on- and offline. Additionally, GPS co-ordinates and images enrich the datasets and make them more accurate.

Another good example is online platforms that help projects aggregate and analyse the collected data in real-time. A centrally accessible web

portal (see screenshot, right) gives the M&E officer in Lesotho a live overview of all monitoring data as it arrives from the districts, enabling her to extract actionable insights. Data visualisations that reflect the project’s indicators allow her to analyse the project’s progress and to identify potential bottlenecks. The ability to export these visuals makes it easier for her to produce regular status reports. And showcasing the live system to district staff can even increase their motivation for reporting reliably and regularly because they see that headquarters is now able to instantly acknowledge their achievements.

### UNDERSTAND THE USER – THEN THINK DIGITAL

These digital solutions benefit the staff in Lesotho because they speak to their needs. For instance, the use of mobile apps makes sense because all project staff have access to smartphones. Other projects, where this may not be the case, would require different solutions. The point is that just by introducing digital solutions, a project will not magically improve its M&E system. This takes more than plug-and-play. Planning officers and project managers need to understand that digital solutions will only be useful if they fit to the project environment and, above all, to the needs and requirements of the people meant to use them.

Even the most sophisticated digital tools for M&E risk failure if introduced based on their technical functions alone. The perspective





Mobile apps can simplify the collection of monitoring data in remote areas.

has to shift from the technology to its users. People are more likely to appreciate digital solutions that meet two basic criteria: firstly, they need to feel useful e.g. because they simplify collecting, analysing or reporting on monitoring data; and secondly, they need to feel easy to use e.g. because they are designed to be intuitive enough not to require lengthy manuals. Therefore, before deciding on any digital solutions, projects need to consult with the people meant to use them and keep them engaged throughout the development process. This gives them the chance to jointly develop a user-friendly digital solution that supports everyone with M&E – and avoid settling on unfitting tools that may end up adding to everyone’s workload rather than relieving it.

## DOING DIGITAL M&E RIGHT

With all this in mind, is there practical advice that projects can follow to apply digital technologies for improving their M&E systems and processes? The answer is yes – but there are no magic recipes. Our experience points to four ingredients that help doing digital M&E right:

1. Assess the digital readiness: Digital tools only take effect if they meet the needs, opportunities and limitations of the environment in which they are to be used. It is therefore important to first assess the digital readiness of the



Online M&E portals can enable projects to aggregate, analyse and extract valuable insights from data in real-time.

project and its context. What skills do people have to operate digital tools? What technical infrastructure is available? What is the mentality towards digital innovation? For instance, a rural project may see potential in introducing a mobile solution to replace paper-based data collection in remote areas. Since people there rely on simple feature phones, the project decides on a text-based system that uses SMS surveys to gather data.

2. Develop agile and with the user: To build useful and intuitive digital solutions, the needs and expectations of the users have to be at the heart of the development. Here an agile, user-centred approach helps. For instance, the rural project may decide to establish a web portal to replace Excel sheets for analysing collected data. This portal is built in several iterations – after each one the M&E staff test it and give feedback. As a result, they feel it is tailored to their needs and appreciate that they can use it early on without complex manuals.
3. Start small and modular: User-centred development also helps projects identify unforeseen requirements and needs with regard to digital solutions. To be able to accommodate these, projects should start with small-scale, modular pilots instead of largely pre-defined

solutions. For instance, the rural project may find that SMS should be used not just for collecting data but also for sharing insights with beneficiaries. The modular pilot makes it easier and less costly for the project to adjust the digital solution – and to bring it to scale eventually.

4. Provide training and support: Even if digital solutions are intuitive and easy to use, projects need to ensure that everyone meant to use them has the digital skills to do so. This means that projects need to invest in training measures followed by continuous support. For instance, the rural project may offer training workshops and identify digital ‘champions’ in its team to support colleagues less digitally adept.

The examples show that there is great potential for projects to reap digital benefits for making their M&E systems more efficient, informative and motivating to use. As long as they stick to one central mantra: Before thinking digital – think of the users.

**Daniel Brumund** is a consultant for digital solutions and communication at Mainlevel Consulting in Berlin/Germany. He advises on strategy and practice of ICTs for M&E and development co-operation.  
Contact: [daniel.brumund@mainlevel.de](mailto:daniel.brumund@mainlevel.de)

# MONITORING RESULTS OF AGRICULTURAL AND FOOD SECURITY PROJECTS: THE INDICATOR CHALLENGE

A results framework, which specifies objectives, indicators and targets, is the basis to monitor progress of development projects. Selecting appropriate indicators and measuring results, however, can be extremely challenging. If indicators are not well selected or data quality is low, their use to support decision-making and to report on results may lead to wrong conclusions. This article provides an overview on criteria to select indicators and explores the main challenges when measuring results in agricultural and food security projects.

By Sarah Holzapfel

Various issues and criteria need to be taken into account when selecting indicators and organising the data collection process. A few of the most important ones are listed in the Box.

- **Direct:** The indicator clearly represents the intended result.
- **Objective:** The indicator is unambiguous about what is being measured.
- **Use of country systems:** The indicator is part of a country-led results framework, and data is provided by country-level M&E systems.
- **Owned:** Partners and stakeholders agree that the indicator is useful.
- **Attributable:** The indicator can be plausibly associated with the intervention.
- **Practical:** Data can be collected easily, on a timely basis and at a reasonable cost.
- **Reliable:** Data is consistent and comparable over time.

There are often trade-offs between the criteria, which have to be carefully weighted. In the following, the most common trade-offs as well as the challenges when defining indicators and measuring results are described.

## DEFINING INDICATORS

Indicators often include a variety of concepts, which have to be defined and clarified. One example is an indicator formulated as “the hectares of land cultivated by smallholders (20 per cent women) under sustainable agricultural practices have increased by 10,000 ha”.

First, the project has to define which practices are understood as sustainable and who is considered a smallholder. Second, it has to be clarified what “20 per cent women” (of 10,000 ha) means. Does the 20 per cent refer to land



Indicators have to be clearly designed. Does “20 per cent women” (of 10,000 ha) refer to land cultivated by female-headed households or to land over which women have primary decision-making rights?

cultivated by female-headed households or to land over which women have primary decision-making rights? Ambiguities often arise with respect to indicators, which measure agricultural productivity, income or food security. If, for example, an increase in productivity by 30 per cent is set as a goal, the project has to specify whether the increase is expected among all households in the intervention area, as a mean among all direct (and indirect) beneficiaries or only among a share of the beneficiaries. These different approaches imply a high difference in the level of ambition, especially if the intervention area is large.

## USE OF COUNTRY SYSTEMS

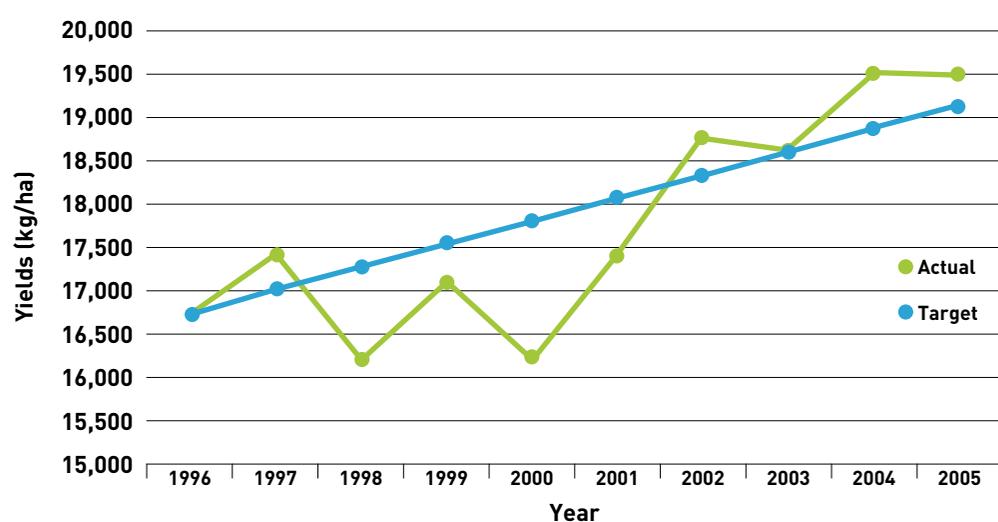
As a best practice and in line with the aid effectiveness agenda, country-led results frameworks and Monitoring & Evaluation (M&E) systems should be used as a common tool among all concerned actors to assess project performance. In practice, donors often introduce their own, parallel monitoring systems and only partly rely on country-level systems. While the objectives of development interventions are usually derived from country-results frameworks, the percentage of indicators that draw on data from country-level M&E systems is much lower. Indicators are often determined by donor priorities and selected with limited involvement of partner countries. However, country systems can best be strengthened by using them. Studies show that parallel monitoring systems create inefficiencies such as parallel and uncoordinated data collection efforts. Harmonising results measurement systems among partner countries and donors could increase data availability and quality.

## DATA AVAILABILITY AND DATA QUALITY

In general, measurement errors for key outcome indicators in agriculture, such as yields, gross margins and smallholder incomes, tend to be particularly high. Among the diverse problems when measuring smallholder crop areas, production and crop value are ill-defined or even non-existent plot boundaries, intercropping, irregular planting density, non-standardised measurement units (e.g. bunch or pieces), a high share of subsistence production and significant post-harvest losses.

Data availability and quality when relying on data from partner countries' own systems are often low, which is one of the main reasons why donors set up parallel M&E systems. In

Detecting a trend in maize yields



Source: Global Donor Platform for Rural Development (2008, p. 28)

many developing countries, continued underinvestment in statistical systems has led to low technical and institutional capacity for producing data from administrative systems or household surveys. As a result, there are no regular, reliable estimates of many key indicators that are commonly used to monitor progress in agricultural and food security development projects. Often, data is only available at the country level but not for the specific area where the project is active. A related problem is that although national statistical agencies, donors, universities and research institutes carry out a high number of surveys, data is often not published or shared. One positive example of how to address this problem is the introduction of the Open Data Policy and Development Data Library by the United States Agency for International Development (USAID).

To be comparable over time, data has to be collected using the same data collection methods and sampling frames. Often, however, a variety of data sets and methods are used within the course of a project. For example, as baseline data, national survey data are used. Later on, a survey is carried out among project beneficiaries using a different sampling frame and questionnaire.

## BASELINE DATA

When collecting baseline data, there is often a trade-off between collecting data for the purpose of planning a project and for monitoring and evaluating it. To plan a project and to set targets, data should be collected as early as possible. At a very early stage of project planning, however, the intervention area might

not yet be narrowed down, and it might not be clear who will be targeted by the project. This creates challenges when results of a project are monitored through follow-up surveys. For example, it might be that baseline data was collected among households who in the vast majority have not participated in the intervention.

One weakness often observed in results frameworks is that projects do not collect baseline data for indicators such as “hectares of land under sustainable agricultural practices” or “number of households who have adopted an innovation”. Instead, it is assumed that the value is zero before the start of the project. This is unlikely to be true and can lead to an overestimation of results.

## MEASURING LONGER-TERM RESULTS AND THE ATTRIBUTION PROBLEM

Development co-operation aims to contribute to long-term goals, such as Sustainable Development Goal (SDG) 2: End hunger, achieve food security and improved nutrition and promote sustainable agriculture. When monitoring development co-operation projects, the aim is therefore to go beyond measuring outputs, i.e. products, capitals goods and services which result from a development intervention, and their short-term effects (e.g. access to and use of services, adoption of innovations). There are, however, two problems when measuring medium- and long-term results.

First, medium- and long-term results are influenced by a variety of external factors, and it is challenging to assess whether observed chang-



A lack of data availability frequently complicates project evaluations. Often, data can only be obtained at country level, and not for the specific area where the project is active.

Photo: Arne Hoel/World Bank

es are a result of the intervention or are due to other influencing factors. For example, agricultural yields or smallholder income are highly variable over time and subject to a variety of external factors such as climate variability and changes in world market prices. Rainfall, or the lack of it, has a particularly strong influence on production, and consequently, yields can vary considerably from one year to the next, especially in non-irrigated conditions, which are prevalent in sub-Saharan Africa. Attribution is practically impossible if country-level data is used to measure progress of an intervention, which is only active at a sub-national level. The challenge of attribution implies that

existing data on the SDG indicators, which is often only available at country level, is in most cases not suitable to measure the results of an intervention.

Second, agricultural development programmes usually take several years until changes at outcome level can be observed because the process of promoting and adopting innovations is time-consuming. Usually, several harvest cycles are needed until changes can be observed. In addition, with the high variability in production, it can take several years until trends in yield levels can be detected (see also Figure on page 27).

## EVALUATION DESIGN

The use of a before-after-comparison among the target group or a comparison of outcomes among programme participants and non-participants is common. However, such methods are not suitable for indicators at medium-term outcome or impact level and may lead to wrong conclusions. A before-after-comparison shows changes among the target group and does not provide information on the causal effect of an intervention. For example, a smallholder's income might have also increased in the absence of a programme – because of his or her own efforts, favourable climatic conditions or high market prices.

It is also problematic to compare outcomes among the treatment group with outcomes among a control group if participation in a programme is based on preferences, decisions and unobserved characteristics (self-selection). Because of self-selection, the control group is unlikely to be statistically identical (on average) to the treatment group. As a result, differences in outcome variables cannot be attributed to the project. It is probable that the two groups would have performed differently even in the absence of the programme.

Causality can be established if rigorous impact evaluation techniques are used, such as randomised assignment, difference-in-differences, and matching. While these are used to evaluate projects, they are only rarely applied for monitoring development projects because they are complex and time-consuming and can be very costly.

If medium-term outcomes and impacts cannot be attributed to a project, are indicators at that level suitable to monitor progress of development co-operation projects? Often, it is advisable to monitor short-term outcomes instead, for example, if studies have shown that the adoption of a specific practice promoted by a project leads to desired changes at outcome and impact level. Moreover, it can generally be expected that if individuals have adopted a practice or technology for a longer time, it has positively affected their lives. Otherwise, disadoption would be the logical consequence.

**Sarah Holzapfel**, PhD, is a research fellow at the German Development Institute in Bonn.

In her research she focuses on how to increase the results orientation of agricultural, rural development and food security interventions in line with the aid effectiveness agenda.

Contact: Sarah.Holzapfel@die-gdi.de

## RECOMMENDATIONS

- **Indicator definition sheets** should be developed for each indicator in results frameworks explaining all relevant information allowing for a proper analysis and interpretation of data (e.g. definition of concepts, data collection method, evaluation design).
- Donors should strengthen their support to **country-level M&E systems**, use them as much as possible, and **harmonise** their results measurement systems.
- Data collected by donors through surveys or as administrative data should be **openly available, accessible, usable and unrestricted**.
- If **baseline data** is collected, surveys should be carried out only after the intervention area is known and sufficient information on who will be targeted is available. Before, secondary data, expert opinions and qualitative methods can be used to plan an intervention and formulate targets.
- **Indicators measuring medium-term outcomes and impacts** should only be used for the purpose of monitoring if they can be plausibly attributed to the intervention (e.g. in case rigorous methods can be applied), and if changes are measurable within the project timeframe. If this is not the case, short-term outcome indicators are more suitable.

# USING STANDARD INDICATORS – OPPORTUNITIES, CHALLENGES AND RISKS

By Sarah Holzapfel, German Development Institute

Standard indicators can be characterised by uniform definition, data collection methods and interpretation. They produce data that can be aggregated (and compared) across interventions, countries or regions, for example. Standard indicators can be distinguished from “custom indicators”, which are formulated to describe specific phenomena or to measure certain changes under unique conditions.

Standard indicators can be formulated at different levels of the results chain. Examples are:

- Output: Number of farmers trained
- Short-term outcome: Land under improved management practices
- Medium-term outcome or impact: Individual Dietary Diversity Score

## OPPORTUNITIES

Standard indicators are used for three main purposes:

1. To align projects towards common goals (planning tool): standard indicators are usually chosen to represent donor priorities. Through mandatory standard indicators, donors aim to ensure that supported projects focus on strategic goals.
2. To compare results across projects (management tool): by using the same standard indicators across projects, donors aim to assess projects' value for money.
3. To report on results (accountability and public relations tool): standard indicators provide a snapshot of aggregate results achieved across countries and interventions.

## CHALLENGES

Unambiguous definitions as well as clear guidelines on data collection methods and instruments are important to ensure that results are comparable and aggregatable. Differences in definitions, methods and instruments can lead to high differences in reported results, which are unrelated to project performance. It is especially challenging to develop guidelines applica-

ble to various project contexts that specify how to measure who benefits from a project. While a project which trains farmers in good agricultural practices can count those who participate in trainings as beneficiaries, it is less clear whom to count as a beneficiary when a project supports the implementation of a national food security strategy through policy advisory work.

## LIMITATIONS AND RISKS

Standard indicators at output and short-term outcome level tend to be very broad in scope in order to be applicable to as many projects as possible. As a result, even if common definitions, methods and instruments are used, results are often not comparable. The IFAD indicator “land under improved management practices”, for example, captures any type of initiatives aimed at promoting sustainable management of natural resources, such as integrated natural resource management practices, agroforestry practices and improved water management practices.

Standard indicators at medium-term outcome or impact level, such as the SDG indicators, are not suitable to measure results of one donor, because they capture changes that are products of the joint efforts of partner countries, donors and other influencing factors.

Standard indicators are a suitable tool to report on aggregate results achieved across interventions and countries. There are risks, howev-

er, when using them to align projects towards common goals and to compare results across projects.

**Alignment:** According to aid effectiveness principles, donors should base their overall support on partner countries' national development strategies. To assess performance of interventions, indicators drawn from country-level results frameworks should be used. If donors use standard indicators to enforce donor priorities, conflicts can arise with the principle of alignment.

**Comparison of results:** Since most standard indicators do not account for qualitative differences in results and context conditions are not considered, their use can lead to two adverse effects. First, standard indicators encourage a focus on quantity instead of quality. For example, an indicator measuring the number of farmers trained does not distinguish between farmers who took part in a two-hour workshop and farmers who participated in a year-long training programme. Second, standard indicators may lead to a focus on low-hanging fruits and a neglect of areas where the per capita costs of delivering results are comparably high. The per capita costs of providing nutrition education to communities in a densely populated part of Cambodia, for example, can be expected to be much lower than in a remote part of the country. Thus, when comparing results across projects based on standard indicators, the context conditions should always be taken into account.

The indicator “number of farmers trained” does not distinguish between farmers who took part in a two-hour workshop and farmers who participated in a year-long training programme.

Photo: Jörg Böhling



## CORPORATE-LEVEL IMPACT MEASUREMENT – IFAD’S EXPERIENCE

For donor countries, corporate-level impact estimates can be crucial for justifying funding to an institution. The International Fund for Agricultural Development (IFAD) has developed a system which allows for monitoring the attributable impact of its entire portfolio on a systematic basis.

By Paul Winters and Alessandra Garbero

Assessing impact requires attribution, which refers to the ability to claim that impact on an indicator of success is the result of a particular investment. Identifying impact entails creating a counterfactual that allows comparison of what has happened as the results of an intervention and what would have happened in the absence of that intervention. As seen in the other articles in this issue, identifying impact at the project level is well understood. Experimental (randomised controlled trials) and non-experimental approaches are becoming widely used to assess impact. These approaches create a counterfactual through a combination of careful data collection and statistical methods which provide confidence that impact estimates are unbiased and thus can be attributed to the intervention.

Attributing corporate-level impact for a development institution, such as the International Fund for Agricultural Development (IFAD), is more complicated and less straightforward. Nevertheless, bodies that govern development institutions are expanding demand for impact estimates that can be attributed to the activities of these institutions. For donor countries, corporate-level impact estimate can be crucial for justifying funding to an institution since it can address taxpayers’ and parliaments’ questions about whether development assistance is effective. For this reason, governing bodies are increasingly asking for Results Measurement Frameworks (RMFs), which lay out the indicators of institutional success and include attributable impact indicators.

### THE AGGREGATION CHALLENGE

Along with the issue of attribution, an additional consideration for corporate-level impact is aggregation. Corporate impact measurement requires having indicators that can be aggregated across a range of interventions as well as a means to add up the overall impacts across those interventions. If every project had the same objective along with the same indicators of success and if every project had an impact evaluation, aggregation would be straightforward; measures of impact could simply be



To increase rural people’s benefits from market participation is one of IFAD’s Strategic Objectives.

Photo: IFAD/G. M. B. Akash

added up. But projects vary in their objectives based on local development needs and country priorities and impact evaluations are costly and cannot be undertaken for every project.

To address the need for attribution and aggregation, IFAD has developed a system for measuring corporate results (see upper Figure on page 31). Based on a theory of change, every IFAD project invests in inputs that are expected to lead to outputs. Provided that beneficiaries of project funding behave in an anticipated manner, this should lead to anticipated outcomes and ultimately impact. For each project, a logical framework (log-frame) is developed with indicators for inputs, outputs, outcomes and impacts. For inputs, outputs and some outcomes attribution is not an issue since the institution knows where its funds go and what their immediate effect is. Data on these indicators is collected based on a monitoring and evaluation (M&E) plan that is developed at the initiation of the project. Since projects necessarily differ because they address distinct development problems in differing contexts, these indicators vary by project. This presents the challenge for aggregation

to the corporate level since it is not possible to aggregate different indicators. But there are sufficient similarities among IFAD projects to allow for similar indicators – what we refer to as Core Indicators. These are mandatory when relevant for all projects and can be aggregated for corporate results reporting.

### INDICATORS REFLECT THE FUND’S STRATEGIC OBJECTIVES

As noted, getting attribution at the impact level is more complicated and costly since it requires substantial data collection efforts on indicators for a treatment (beneficiaries) and control (counterfactual) group. It is therefore difficult to justify impact assessments for all projects. For this reason, the impact level measurement is done in 15 per cent of projects. Of course, this creates a problem in that corporate-level impact should be an estimate of total impact, not just 15 per cent of the portfolio. For this reason, IFAD has devised an aggregation “methodology”. The first step in this process is the identification of the indicators and targets to measure in order to reflect corpo-

rate success. The overarching goal of IFAD's Strategic Framework 2016–2025 is to invest in rural people to enable them to overcome poverty and achieve food security through remunerative, sustainable and resilient livelihoods. To achieve this goal, the Fund identifies three strategic objectives: SO1 – increase rural people's productive capacities; SO2 – increase rural people's benefits from market participation; and SO3 – strengthen the environmental sustainability and climate resilience of rural people's economic activities. In accordance with this strategy, the following impact indicators are used in the RMF and defined in such a way that they can be aggregated across projects:

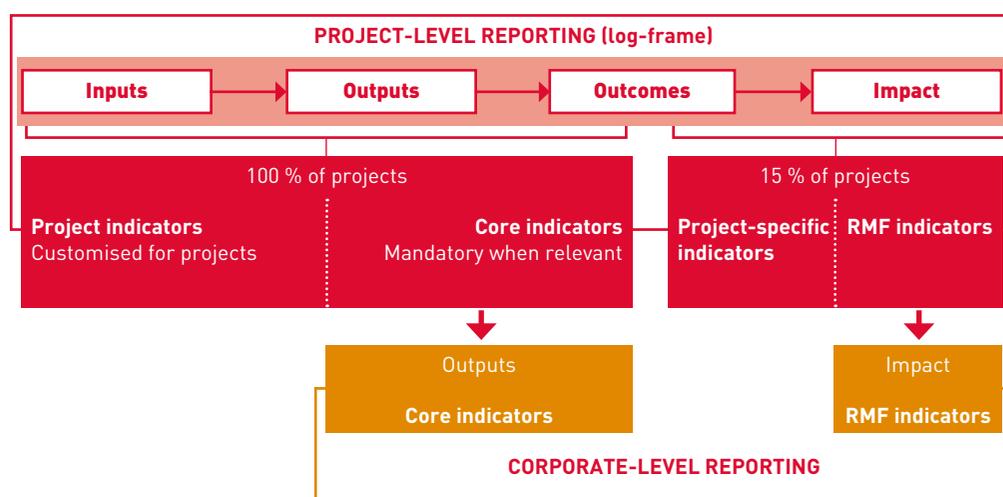
1. Number of people experiencing economic mobility (goal)
2. Number of people with improved production (SO1)
3. Number of people with improved market access (SO2)
4. Number of people with greater resilience (SO3)

As with Core Indicators not all indicators for an impact evaluation are RMF indicators since projects may include other context-relevant indicators.

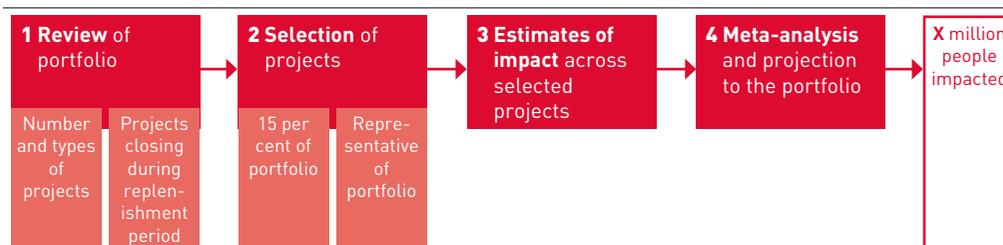
While impact evaluations provide estimates of the average project impact, these indicators are reported in terms of a number of people receiving a benefit. The lower Figure summarises the approach to aggregating from the 15 per cent of projects with impact evaluations to corporate reporting of the number of people benefiting.

Step one in aggregating impact is to understand the portfolio of investments. This includes investment projects that are being completed during a particular period of interest, which for IFAD is the three year replenishment cycles (e.g. IFAD10: 2016–2018). This means having a sense not only of the amount of investment, but also of the types of investment. The second step is to select the 15 per cent of projects suitable for an impact evaluation. These need to be selected to represent the types of projects in the portfolio, so that any aggregation reflects the whole portfolio. The third step is to conduct the impact evaluations of individual selected projects and estimate project-specific impact. This provides the average impact of each project on project beneficiaries. The final step is to take these estimates and use a methodology, for instance meta-analysis, to obtain a global average impact estimate, and employ a set of

### IFAD corporate approach to reporting



### IFAD steps for assessing impact



Source: IFAD

assumptions that allow one to project such estimate to the overall portfolio. This requires a clear understanding of the portfolio and a set of assumptions regarding how average impacts translate into number of people benefiting. The details of this are beyond the scope of this article, but the logic should be clear. The basis for the estimates are impact evaluations that allow for attribution, carefully selected aggregable indicators and a clear understanding of the portfolio that allows for aggregation.

### CONSIDERING THE ENTIRE PORTFOLIO

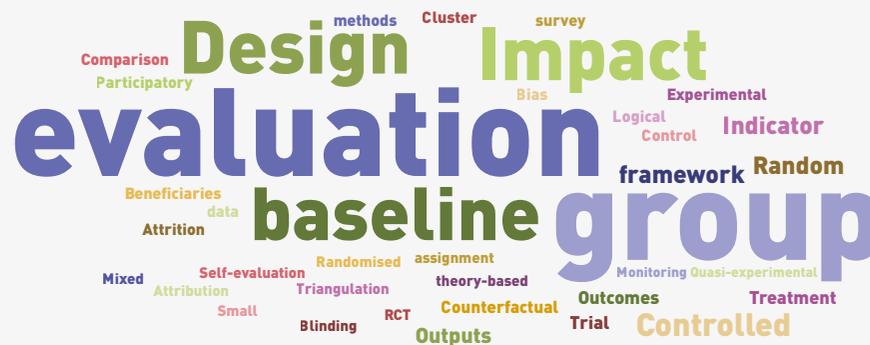
While relying on a number of assumptions, this system enables reporting corporate-level impact measurement, something that has not been done by many development institutions. Although this is important for accountability and reporting to governing bodies, it allows for a systematic review of the portfolio and for learning on how to improve projects. Standard practice for impact evaluation is to select projects based on the ability to learn lessons but also reflecting the feasibility of assessing impact and the interests of researchers. A corporate-level approach ensures that an entire portfolio is considered and that lessons learned on

whether an institution is doing things right as well as doing the right things can be gathered in an efficient manner.

Of course, there are numerous challenges in doing this, and the system needs to be reconsidered and improved. Along with methodological issues, it requires technical expertise and significant resources that could be used elsewhere. This means that the value of the impact evaluations needs to be sufficient to justify the costs. The entire process must draw lessons that feed into future decisions on investments and help improve approaches to development.

**Paul Winters** is Associate Vice President, a.i., Strategy and Knowledge Department and Director, Research and Impact Assessment Division at the International Fund for Agricultural Development (IFAD). Previously, he was a Professor of Economics at American University in Washington, DC. Contact: p.winters@ifad.org

**Alessandra Garbero** is a Senior Econometrician in the Research and Impact Assessment Division at IFAD. She previously worked for the International Institute for Applied Systems Analysis, on population forecasting, demographic modelling and vulnerability to climate change. Contact: a.garbero@ifad.org



## A BRIEF GLOSSARY OF IMPACT EVALUATION TERMS

### Attribution

The extent to which the observed change in outcome is the result of the intervention, having allowed for all other factors which may also affect the outcome(s) of interest.

### Attrition

Either the drop out of participants from the treatment group during the intervention, or failure to collect data from a unit in subsequent rounds of a panel data survey. Either form of attrition can result in biased impact estimates.

### Baseline survey/baseline data

A survey to collect data prior to the start of the intervention. Baseline data are necessary to conduct double difference analysis, and should be collected from both treatment and comparison groups.

### Beneficiaries

The individuals, groups, or organisations, whether targeted or not, that benefit, directly or indirectly, from the development intervention.

### Bias

The extent to which the estimate of impact differs from the true value as result of problems in the evaluation or sample design (i.e. not due to sampling error).

### Blinding

A process of concealing which subjects are in the treatment group and which are in the comparison group, which is single-blinding. In a double-blinded approach neither the subjects nor those conducting the trial know who is in which group, and in a triple-blinded trial, those analysing the data do not know which group is which.

### Cluster evaluation

An evaluation of a set of related activities, projects and/or programmes.

### Comparison group

A group of individuals whose characteristics are similar to those of the treatment groups (or participants) but who do not receive the intervention. Under trial conditions in which the evaluator can ensure that no confounding factors affect the comparison group, it is called a control group.

### Control group

A special case of the comparison group in which the evaluator can control the environment and thus limit confounding factors.

### Counterfactual

The state of the world in the absence of the intervention. For most impact evaluations the counterfactual is the value of the outcome for the treatment group in the absence of the intervention. However, studies should also pay attention to unintended outcomes, including effects on non-beneficiaries.

### Impact

How an intervention alters the state of the world. Impact evaluations typically focus on the effect of the intervention on the outcome for the beneficiary population.

### Impact evaluation

A study of the attribution of changes in the outcome to the intervention. Impact evaluations have either an experimental or quasi-experimental design.

### Indicator

A quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor.

### Large n impact evaluation

Studies applying statistical means to construct a counterfactual, which requires a sufficiently large sample size (n) to ensure statistical power.

### Logical framework (Logframe)

A management tool used to improve the design of interventions, most often at the project level. It involves identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators and the assumptions or risks that may influence success and failure. It thus facilitates planning, execution and evaluation of a development intervention.

### Mixed methods

The use of both quantitative and qualitative methods in an impact evaluation design. Sometimes called Q-squared or Q2.

### Monitoring

A continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds.

### Outcome(s)

A variable, or variables, which measure the impact of the intervention.

### Outputs

The products, capital goods and services which result from a development intervention; may also include changes resulting from the intervention which are relevant to the achievement of outcomes.

### Participatory evaluation

Evaluation method in which representatives of agencies and stakeholders (including beneficiaries) work together in designing, carrying out and interpreting an evaluation.

### Quasi-experimental design

Impact evaluation designs used to determine impact in the absence of a control group from an experimental design. Many quasi-experimental methods, e.g. propensity score matching and regression discontinuity design, create a comparison group using statistical procedures. The intention is to ensure that the characteristics of the treatment and comparison groups are identical in all respects, other than the intervention, as would be the case from an experimental design. Other, regression-based approaches, have an implicit counterfactual, controlling for selection bias and other confounding factors through statistical procedures.

### Random assignment

An intervention design in which members of the eligible population are assigned at random to either the treatment group or the control group (i.e. random assignment). That is, whether someone is in the treatment or control group is solely a matter of chance, and not a function of any of their characteristics (either observed or unobserved).

### Randomised controlled trial (RCT) / Experimental design

An impact evaluation design in which random assignment has been used to allocate the intervention amongst members of the eligible population that is meant to ensure that there is no correlation between participant characteristics and the outcome. Differences in outcome between the treatment and control group can be fully attributed to the intervention, i.e. there is no selection bias.

### Self-evaluation

An evaluation by those who are entrusted with the design and delivery of a development intervention.

### Small n impact evaluation

The set of best available methods when n is too small to apply statistical approaches to constructing a counterfactual.

### Theory-based impact evaluation

A study design which combines a counterfactual analysis of impact with an analysis of the causal chain, which mostly draws on factual analysis.

### Theory of change

Laying out the underlying causal chain linking inputs, activities, outputs and outcomes, and identifying the assumptions required to hold if the intervention is to be successful. A theory of change is the starting point for theory-based impact evaluations.

### Treatment group

The group of people, firms, facilities or whatever who receive the intervention. Also called participants.

### Triangulation

The use of three or more theories, sources or types of information, or types of analysis to verify and substantiate an assessment. By combining multiple data-sources, methods, analyses or theories, evaluators seek to overcome the bias that comes from single informants, single methods, single observer or single theory studies.

Source: OECD (2009): *Glossary of Key Terms in Evaluation and Results Based Management*; International Initiative for Impact Evaluation (2012): *3ie impact evaluation glossary*.



# LAND DEGRADATION NEUTRALITY – A NEW IMPETUS FOR ADDRESSING THE DEGRADATION OF LAND AND SOILS

With the target on Land Degradation Neutrality (LDN) the Sustainable Development Goals have created new political momentum to address land degradation. The UN Convention to Combat Desertification adopted LDN as its central objective and supports countries in defining national targets to avoid degradation and restore degraded land. This provides an opportunity to advance action against land and soil degradation on national agendas, reform respective policy instruments and mobilise additional funding.

By Alexander Erlewein and Antje Hecheltjen

Land degradation is an increasingly severe global environmental and development problem. Each year an additional area of 12 million hectares of agricultural land get degraded and soil erosion amounts to an estimated 24 billion tons (3 tons per capita). The manifestations of land degradation are diverse and context-specific, but are always characterised by the degradation of soil, vegetation and/or water resources – predominantly through unsustainable forms of land use. This results in a massive loss of ecosystem services, with the reduction in agricultural productivity being a main concern. The increasing degradation of land and soils is therefore a major threat to food security and the resilience of rural communities while contributing to climate change and biodiversity loss. In total, the estimated annual costs of land degradation world-wide amount to 400 billion US dollars.

## A NEGLECTED PROBLEM

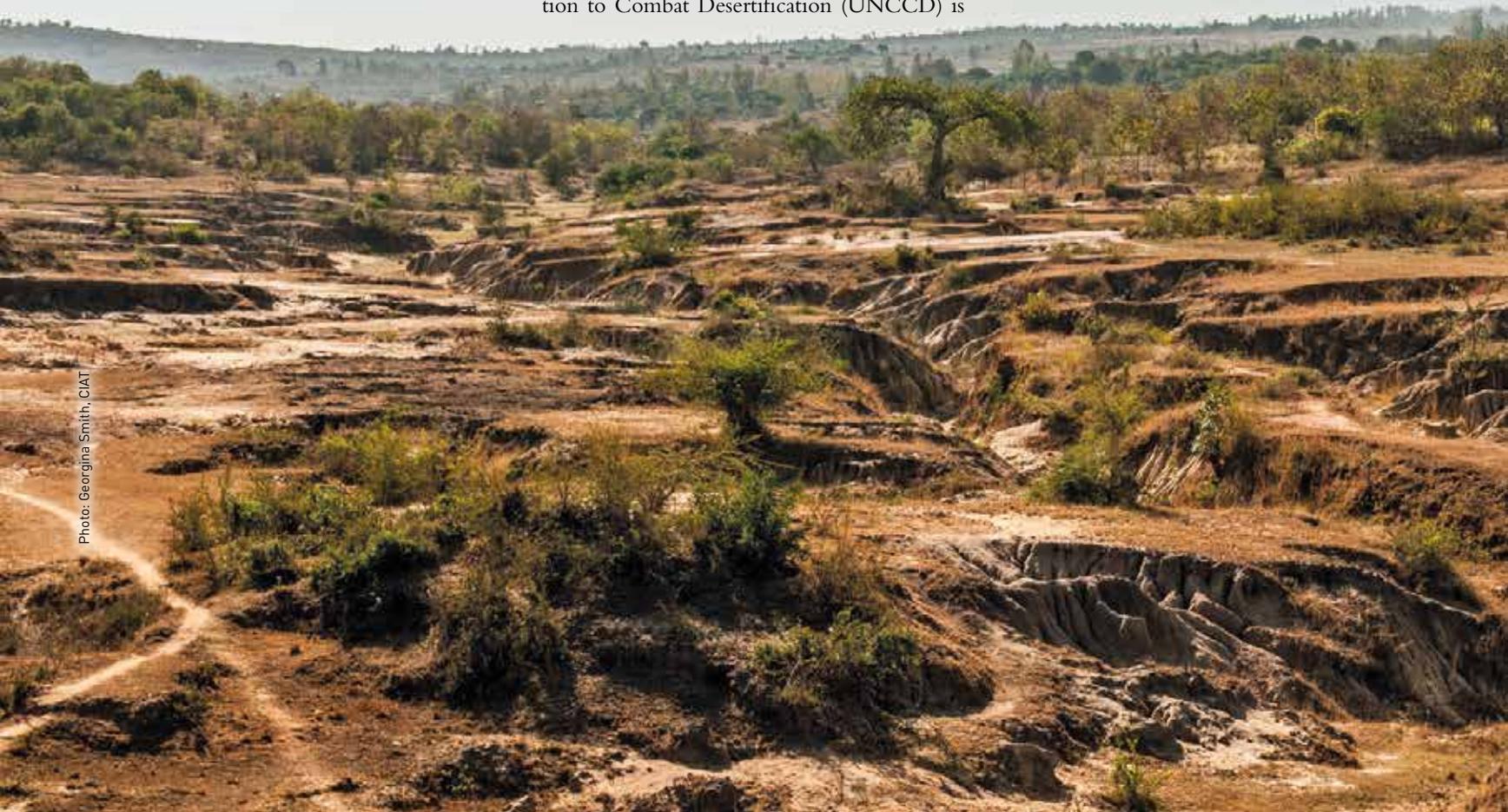
Irrespective of the fundamental challenge that land degradation poses to rural development in many parts of the world, awareness among the public and policy-makers is generally low. Soil continues to be a neglected resource whose degradation usually takes place slowly and only becomes visible at a late stage. Moreover, land and soil are often perceived as private property rather than public goods, while political responsibility cuts across the agricultural and environmental ministries. Whereas the agricultural sector tends to perceive soil fertility as a mere function of input supply, the environmental sector largely reduces land degradation to land cover change, with a focus on deforestation.

At the international level, the UN Convention to Combat Desertification (UNCCD) is

the only legally binding international agreement to address land and soil degradation. Despite being one of the three Rio Conventions, the actual influence of the UNCCD has been limited in the past. This may be explained by its formally restricted mandate on land degradation in drylands, i.e. desertification, and a regional focus on Africa. More importantly though, the convention lacks a clear and quantifiable target.

## AWARENESS OF LAND DEGRADATION IS GROWING

However, attention has been growing over the last few years. A number of international initiatives (such as the Global Soil Week, the Food and Agriculture Organization's Global Soil Partnership or the Economics of



Land Degradation Initiative) as well as several high-ranking scientific assessments have raised the awareness of decision-makers. A particularly important step was the integration of land and soil degradation into Agenda 2030. SDG target 15.3 stipulates to “combat desertification, restore degraded land and soil, including land affected by desertification, drought and floods, and strive to achieve a land degradation-neutral world by 2030”. This target started to gain significant political momentum when the 12<sup>th</sup> Conference of the Parties to the UNCCD decided to make LDN the central objective for the convention in 2015.

### WHAT IS LAND DEGRADATION NEUTRALITY?

Land degradation neutrality (LDN) is first of all an **aspirational target**. Similar to the role of the two degree target in global climate policy, LDN serves as a common overarching goal to address a global environmental problem, giving orientation to the UNCCD process and providing a joint vision for the often fragmented strategies to address land degradation. Constituting benchmarks to which countries and the international community can be held accountable for politically, targets aim to eventually spur action.

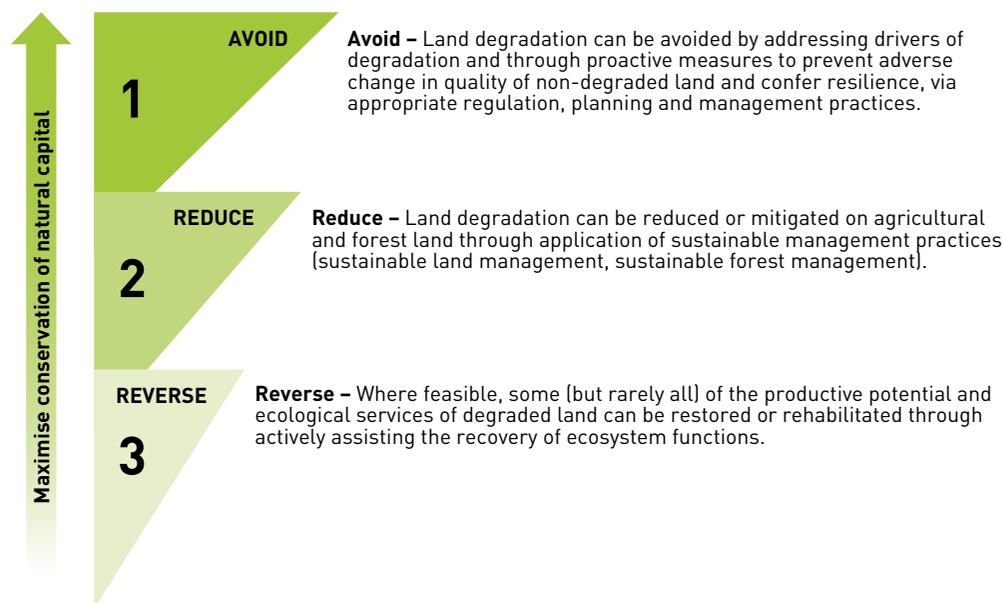
Secondly, LDN is a **concept**, defined as “a state whereby the amount and quality of land resources necessary to support ecosystem functions [...] remain stable or increase”. Thus, LDN is achieved if there is no net loss (or a gain) of land resources compared to a baseline (e.g. 2015). Such a balance can be achieved through avoiding, reducing and reversing land degradation. The LDN response hierarchy (see Figure) underlines the need to prioritise the avoidance of land degradation while making use of the large areas of degraded land that holds potential for restoration.

Thirdly, LDN is a **monitoring approach** that allows for tracking progress on the implementation of LDN targets (see Box on page 35).

### TRANSLATING A GLOBAL GOAL INTO NATIONAL AMBITIONS: LDN TARGET SETTING

Based on these considerations, the UNCCD has started a LDN target-setting programme that aims to bring the global goal of LDN down to the country level. Since its start in 2015, 115 countries have joined the programme with the objective to formulate vol-

#### The LDN response hierarchy



Source: Scientific Conceptual Framework for Land Degradation Neutrality. A Report of the Science-Policy Interface. UNCCD, 2017.

untary national targets on reducing land degradation and rehabilitating degraded land. So far, 60 countries have set LDN targets, some with rather general goals for achieving a state of no net loss in 2030 or at an earlier or later date and others with specific quantitative targets. For example, Senegal aims to annually improve 5 per cent of the land under degradation until 2030, while Namibia has committed to reduce bush encroachment on 1.9 million hectares by 2040 (see also article on pages 36–37).

Such ambitions may overlap with other already existing ones for forest and landscape restoration (AFR 100, 20x20 etc.) or climate action (Nationally Determined Contributions). The LDN target-setting programme explicitly encourages such linkages as they provide leverage for political and financial support.

### DIFFERENT PATHWAYS FOR IMPLEMENTING LDN

With a growing number of countries having set LDN targets, the challenge increasingly becomes one of implementation. Bold targets need to be translated into policies and projects that actually improve the way land is managed on the ground. There are two main approaches for countries to implement LDN targets. One is policy-oriented and aims to mainstream LDN into existing land use regulation. In this respect, the LDN mitigation hierarchy (avoid, reduce, reverse) can serve as a guiding principle for land use planning. The other implementa-

tion pathway is more project-oriented aiming at the roll-out of additional programmes for sustainable land management and rehabilitation in degradation hotspots. Such strategies are more concrete and help to achieve tangible results, although they are also more limited in scope and may not address the need for policy reform. A combination of policy and project-oriented approaches is clearly best suited to successfully implement LDN targets.

Obviously, many existing programmes and projects already contribute to achieving national LDN targets, particularly those working on landscape restoration, agroforestry, watershed management, soil rehabilitation or erosion control. They should be made visible and may help to identify scalable best practices and increase the ambitiousness of LDN targets. An LDN target a country has set can be an opportunity to advance necessary reforms of policy and planning instruments as well as to foster the often lacking co-operation between agricultural and environmental authorities.

### PROMISING EXAMPLES AND OBSTACLES

Experience from countries following an LDN approach is diverse. The number of countries participating in the LDN target setting programme, now including global players such as China, India and Brazil, clearly exceeds expectations. However, the political relevance of LDN targets varies. While some countries of-

## MONITORING LAND DEGRADATION NEUTRALITY

What is the status of land degradation and land rehabilitation worldwide? Has it improved or worsened? Where are degradation hotspots and anticipated losses located? Where should a country plan measures to avoid, reduce, or reverse land degradation?

To answer these questions, decision-makers need spatial data. Freely available satellite imagery in combination with in-situ observations provide reliable data with global coverage and high repetition rates, enabling retrospective and current analysis. Most countries, however, are not yet exploiting the potential of data for monitoring, reporting and planning purposes. Therefore, the UNCCD, together with its partners, supports all country parties by providing readily processed data to each country as well as free tools for visualising and processing their own data and by building capacities through regional workshops.

### Data for reporting

In 2018, parties to the UNCCD will report for the first time on the following three agreed quantitative indicators in a uniform approach that allows the spatially explicit estimation of changes in land degradation and restoration (see Figure):

- Land productivity
- Land cover change
- Soil organic carbon (SOC)

The three indicators are a minimal consensus to quantify land degradation. From a scientific point of view, more sophisticated approaches exist to estimate land degradation. For monitoring, however, an approach was needed that is (a) accepted by all country parties and (b) operational at global level. That this consensus was reached and that monitoring is now operational is a huge step forward for the Convention. Countries are explicitly invited to add country-specific indicators and national data where needed.

Their choices on data will always be a compromise between (a) global comparability and (b) national relevance to inform national decision-making on resource management and land use planning.

The same indicators used for the UNCCD reporting process will also serve for reporting on SDG 15.3 on LDN, avoiding duplication of efforts; however, some countries are still struggling aligning the two reporting processes. The required information flow between UNCCD national focal points and national statistical offices as well as the acceptance of geospatial data by statisticians can be challenging.

ficially endorsed their targets, others consider them to be more informal. Targets formulated by environment ministries without the participation of their agricultural counterparts are a recurrent obstacle.

LDN explicitly calls for cross-sectoral collaboration, and most targets can only be implemented by improving agricultural practices, incentives and advisory services. A promising example in this regard is Costa Rica, which already passed a directive that requires all ag-

ricultural and environmental policies and plans to integrate LDN. In Benin, the LDN process has triggered a continuous inter-ministerial dialogue on sustainable soil management, supported by a bilateral soil rehabilitation project that provides best-practice examples.

Many countries follow a pragmatic approach and use their recently set LDN targets as an argument to access additional funding for project proposals. In fact, the Global Environmental Facility (GEF) recently made LDN the cor-

nerstone of its focal area on land degradation and significantly increased respective funding volumes. A number of GEF financed LDN projects have already been approved (e.g. in Georgia, Lebanon, Namibia and Turkey), usually combining on-the-ground implementation with activities to integrate LDN elements in policy processes. Another emerging funding opportunity is the recently launched LDN Fund that aims to mobilise private sector investments for profit-oriented projects with high environmental and social benefits.

## FRAMEWORK FOR MONITORING AND REPORTING

### SDG Indicator 15.3.1

Proportion of land that is degraded over total land area



### Sub-Indicators UNCCD (CBD, UNFCCC) Reporting Mechanisms



### Data from multiple sources



Source: UNCCD

### Data for decision-making

Besides reporting, the data will serve as a basis for policy-making by informing land use planning with the aim to optimise the location of interventions and the type of interventions. Namibia, for example, used national data on the three indicators plus bush encroachment as a country-specific indicator to inform the integrated regional land use planning (IRLUP) process. Data on bush encroachment was especially useful, while the interpretation of SOC data proved difficult. To ensure maximum exploitation of the potential of the data for planning processes, capacity building should target both reporting officers and technical staff for land use planning.

### Synergies and inter-sectoral co-operation

The three indicators – especially land cover change – are also relevant for reporting and decision-support for other SDGs and international agreements and target systems such as AFR100, the Sendai framework for disaster risk reduction, or Nationally Determined Contributions (NDCs). Inter-sectoral co-operation is thus of utmost importance to maximising synergies and avoiding duplication of efforts for monitoring, assessment, planning and implementation to move towards a land degradation neutral world.

## SUMMING UP...

Undoubtedly, LDN has given the global agenda on land degradation a new boost. It provides a joint vision and monitoring approach for the fragmented policy field of land and soil degradation and encourages a significant number of countries to put these topics higher on the political agenda. This offers an opportunity to advance action against land and soil degradation, reform respective policies and mobilise additional funding. A key aspect is the flexibility of the LDN process, which gives countries sufficient freedom to choose their targets, implementation paths and monitoring indicators according to their specific circumstances.

**Dr Alexander Erlewein** is advisor on land and soil degradation at GIZ. He previously worked for the UNCCD secretariat.

**Antje Hecheltjen** is advisor on land degradation and remote sensing at GIZ. She previously worked for UN-SPIDER and as a consultant for UNCCD. Contact: ccd-pr@giz.de



Participative mapping is key to land degradation neutrality.

Photo: Klaus Ackermann

## MEASURING LAND DEGRADATION NEEDS TO BE DONE FROM THE GROUND UP

Bush encroachment is a matter of growing concern in the semi-arid regions of Namibia. Locally adapted, long-term solutions for land restoration must be elaborated. It is crucial that local scientists are empowered to map and monitor the degradation process themselves, our authors maintain.

By John Yumbya Mutua and Ravic Nijbroek

The Herero people of Namibia are resilient. After migrating from the Great Lakes of East Africa as herdsmen, they faced invasion by armed migrants and survived bitter warfare throughout the 19<sup>th</sup> century. But today, in the remote, semi-arid Otjozondjupa Region, where they have their communal home, the Herero people face another threat – felt this time under their feet. The grassland their animals rely on is being replaced with dense, thorny bushes, a process called “bush encroachment”.

Commercial livestock farmers in the country face the same threat. The beef industry is a major economic sector in Namibia accounting for 68.4 million US dollars of the country’s export revenue annually. But bush encroachment is causing revenues to drop. This form of land degradation is now recognised as the fourth sub-national indicator in Namibia under the

UN Convention to Combat Desertification (UNCCD) process of Land Degradation Neutrality, which is setting a process in motion to map, restore and monitor degraded land on a global scale.

### A FAST SPREADING PROBLEM

Our research shows rapid encroachment of high and low density bushes creeping into grassland areas (see Figure on page 37). This has important consequences for the land degradation process for a number of reasons. As an example, our preliminary findings show that significant amounts of carbon stored in the soil are lost after grasslands have been converted to bushland. It takes decades before bushes can sequester – or store – more carbon in the soil again. In addition, bush encroachment is

thirsty: the bushland roots reach much deeper into the soil than grasslands, sucking up valuable water resources, which in turn makes less water available for grasses and changes the local ecology.

While bushes are a natural feature of Otjozondjupa Region, it is not clear what is causing the rapid increase in bush density, and as yet, no cost-effective method has been identified to stop the encroachment. Some evidence suggests that livestock herding may itself be a driver, since bushes may be replacing grasses that have been overgrazed. Some entrepreneurs are trying to turn the problem into an opportunity by producing charcoal for the European market. What is certain is that local solutions for people like the Herero and other marginalised groups will need locally relevant solutions on the ground.

## BUILDING CAPACITY ON THE GROUND

In 2016, we embarked on assessing the extent of land degradation in Otjondzupa Region in Namibia, and partnered with a team from the country's Ministry of Environment and Tourism (MET) – some of whom represent the government at the UNCCD – to establish a baseline to measure Soil Organic Carbon (SOC) and Bush Encroachment (BE). In order to develop the baselines to track encroachment, we first worked out baselines to track Land Use and Cover Change (LUCC) and Net Primary Productivity (NPP), which helps us distinguish bushes from grasses remotely. This was done by analysing different climatic, vegetative and topographic variables and data – which had not been carried out at such a scale and resolution before. Critically, all four of these variables are also the main indicators used by the government to report progress on Land Degradation Neutrality to the UNCCD.

While the CIAT team is based in Nairobi, Kenya, and our collaborators, World Soil Information (International Soil Reference and Information Centre – ISRIC) operating from Wageningen University in the Netherlands, we knew from the start that the success of land restoration efforts would depend on empowering local scientists to map and monitor the degradation and restoration processes

and, more importantly, that methodologies needed to be adapted locally. Namibia's model of Integrated Regional Land Use Planning (IRLUP) is based on stakeholder engagement with local communities. It is up to these communities to do the actual land restoration. If the activities involved are not accepted by the communities, they will likely fail. So it is important for them to be locally relevant and context-specific. Twenty land resources management professionals were trained in remote sensing, geographic information systems (GIS) and digital soil mapping (DSM) technologies using open source software; this enables participants to reproduce LDN baselines without additional software costs in the future.

## THE NEXT STEPS

As part of a series of trainings, teams from different ministries and universities now lead the effort in collecting soil samples, identifying plant species, and digital soil mapping. We are now in the final stages of validating the LDN baseline results for a different region, Omusati. Enthusiasm for further collaboration is high, and applications of the newly acquired skills are limitless.

All regions are now required by law to complete IRLUP. As part of the process, important data, e.g. the land degradation baselines, are to

Namibia has established a national de-bushing programme which promotes the large-scale expansion of effective activities to fight bush encroachment. The programme is supported by public and private sector stakeholders. GIZ backs both this programme and IRLUP.

be included in participatory meetings with local communities such as the Herero. Yet the biggest achievement of our work may still be to come. Once communities like the Herero can use locally produced maps to co-develop sustainable approaches for managing bush encroachment, increasing soil carbon, water resources and improving livelihoods, we expect to see ground-breaking, locally relevant solutions.

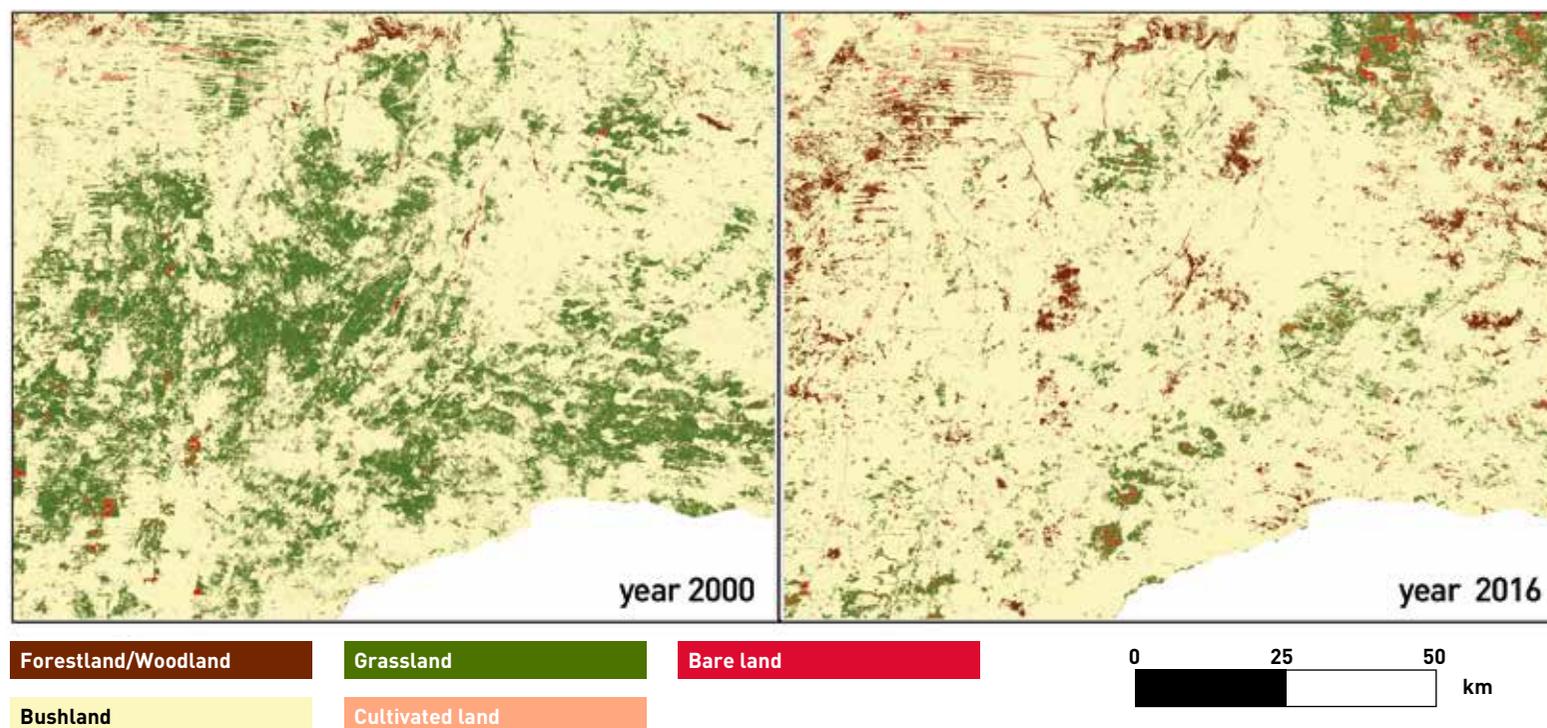
**John Yumbya Mutua**, MSc, works as a Geographic Information System Analyst in the Soils and Landscapes for Sustainability (SoiLS) Program at the International Center for Tropical Agriculture (CIAT) in Nairobi, Kenya.

Contact: [j.y.mutua@cgiar.org](mailto:j.y.mutua@cgiar.org)

**Ravic Nijbroek**, PhD, works as a Landscape Scientist in the Soils and Landscapes for Sustainability Program at CIAT and is based in Nairobi, Kenya.

Contact: [r.nijbroek@cgiar.org](mailto:r.nijbroek@cgiar.org)

## Land cover change in the study area



## A SMALL EFFORT AND A BIG IMPACT – FROM PEDESTRIAN TO TRACTORABLE BRIDGES

With a relatively simple improvement of the traditional Nepalese trail bridges, farmers in the Laotian province of Luang Namtha can now manage their fields more efficiently and bring their produce to the market. A South-South co-operation scheme in technology transfer.

By Niraj Acharya and Bindu Khambu

Physical access to basic services and economic opportunities is a key to poverty reduction of rural people. However, given the huge investments it incurs, providing road transport facilities is not easy in many rural settlements of under-developed and developing countries. Therefore, a trail-based transport system will continue to play a major role in improving rural accessibility for some years to come. In this regard, trail bridges have been a vital element of the system. A trail bridge is a structure built in hilly cum mountainous areas or the plains across rivers and streams along the foot trail, in order to facilitate the safe crossing of pedestrians and animals with or without loads.

### A HIGHLY APPROPRIATE LOCAL TECHNOLOGY

Nepal has been the evolving ground for trail bridges since the 1950s. So far, about 7,000 trail bridges have been built in the country. On average, a further 350 trail bridges are added annually to this total. According to a Post Bridge Building Assessment conducted in 2015, average daily traffic per bridge is 208, which amounts to about

1.4 million people using the bridges every day for various reasons. Trail Bridges have multiple uses: from visiting market centres to going to schools and health service centres, attending social functions, doing agricultural works and performing household chores.

A trail bridge is a highly appropriate local technology, both construction materials and skills being sourced locally, thereby making it a low cost and easily implemented form of rural infrastructure. In Nepal, communities have been building trail bridges (wooden log bridges and chain bridges) across Himalayan rivers using indigenous technology for centuries. Engineered trail bridge building started only around early 1900s. Bridges were assembled in Scotland, parcelled to Nepal and erected in a few strategically important market centres or main trails.

However, systematic and planned trail bridge construction only started around the mid-1960s, when the Government of Nepal established the Suspension Bridge Division. Since then, the technology has been greatly improved and adapted to the local needs. Now,

Nepal provides technical assistance to other countries, especially Southern nations, to transfer and adapt the trail bridge technology in their context.

### INNOVATION FROM PEDESTRIAN TO TRACTORABLE BRIDGES

Nalae and Long districts of Luang Namtha province in Laos are predominantly agrarian areas. Despite heavy reliance on labour-based farming, the use of tractors in farming is a common practice in these areas. However, without proper bridges over the Namyang and Namkha Rivers, farmers could not take their tractors to land on the other sides of the rivers, and had to rely on manual labour, thereby reducing the efficiency of farming. A tractorable bridge is useful to the rural people to transport their agricultural products directly from farm land to market or home. Helvetas Swiss Intercooperation in Laos realised this need and approached Helvetas Swiss Intercooperation Nepal's south-south co-operation team to pilot a couple of river crossings that can accommodate power tillers (hereafter referred to as tractors).





The tractorable Skan bridge in Laos.

Photo: Bindu Khambu

The organisation's bridge team designed a tractorable bridge, which is an improvisation of the trail bridge technology developed in Nepal.

Pedestrian trail bridges are designed for maximum 1.2 m walkway width (see Box). For tractorable bridges, the walkway width was increased to 1.7 m in order to accommodate a 1.45 m wide tractor. Increasing the walkway width has an implication on the design load of the bridge. The self-weight of a standard tractor and the maximum loads it can carry were considered for design. Similarly, the dead load was also increased according to the increase in size of bridge components (e.g. walkway, cross beam, etc.). Suspended and Suspension bridges are flexible structures, hence they deform in the direction of load significantly. In order to control the vibration of the bridge while moving loaded tractors over them, a wind guy system was introduced irrespective of the bridge span. In a pedestrian bridge, a wind guy is generally only provided for bridges over 120 m in length. It enhances the lateral stability of the bridge. The construction of the bridges was supported by the Poverty Reduction Fund (financed by the World Bank and Swiss Agency for Development and Cooperation) and outsourced to contractors. Since these were the first bridges of their kind in Laos, Helvetas provided technical assistance to the Fund for quality control and monitoring of the bridge construction.

### A COST-EFFECTIVE SOLUTION WITH NUMEROUS SOCIAL BENEFITS

Chapee (see photo on page 38) and Skan (see top photo) are the two bridges built with this technology. In both places, crossing the rivers was unsafe prior to their construction. There was no bridge in Chapee, and people used to risk their life when crossing the river, especially during floods. They even had to spend

## PEDESTRIAN TRAIL BRIDGE TYPES

### Steel truss bridge

A pedestrian steel truss bridge is built over rivers either in the hills cum mountains or the plains, with a span limited to 32 m. It is a firm steel structure that is manufactured in the workshop and assembled at the site.

### Suspended bridge

A suspended bridge is built on sufficiently high river valley that allows required free board (minimum 5 m) from the highest flood level. It is economical, simple to design and construct in comparison to other bridge types. It has a downward sagging steel walkway fitted directly on the main cables (load bearing cables) which are supported by a network of suspenders attached to the handrail cables. Since it is a cable bridge, it vibrates when walked upon. A wind guy system is provided to control the excess sway if the span of the bridge exceeds 120 m.

### Suspension bridge

The suspension bridge is designed for flatter terrain so as to attain minimum required free board. It has an upward cambering steel walkway connected to the main cables by the steel suspenders. Steel towers are provided at both ends to hold the main cables which are then anchored to the foundation blocks. This type of bridge is costly, and complex in design and construction in comparison to suspended bridge. It demands higher technical skill to design and construct. As in suspended bridge, a wind guy system is provided to control the excess sway only if the span of the bridge exceeds 120 m.

nights in the forest waiting for the flood to subside. Whilst in Skan, there used to be a temporary wooden log bridge. With the construction of these bridges, local community access to schools, market centres, and agriculture farms was enhanced. The local people are now able to transport their agro-products directly to the market on tractors. Furthermore, motorcycle is a popular mode of transport in these areas and these bridges allow two-way motorcycle traffic, hence significantly reducing travel time. This has proved crucial, especially in the situation of health emergencies.

These pilot bridges are highly appreciated by the local people as well as the government officials. The simple improvisation of the trail bridge technology has proved to be a promising solution for rural farmers. It could be a new cost-effective transport approach in many areas in south and south-east Asia where con-



A pedestrian truss bridge in Nepal.

Photo: Pasang Sherpa



A pedestrian suspended bridge in Nepal.

Photo: Niraj Acharya



A pedestrian suspension bridge in Ethiopia.

Photo: Seyoum Debebe

struction of a motorable bridge is financially not viable. However, care has to be taken that these bridges really are only used by the vehicles that they have been designed for.

**Niraj Acharya** is a rural development professional and currently works as Water and Infrastructure Coordinator for HELVETAS Swiss Intercooperation based in Nepal.

Contact: niraj.acharya@helvetas.org

**Bindu Khambu** is a civil engineer who has been in the Trail Bridge sector since 1981. At present, he is working as a trail bridge expert in the South-South Cooperation Unit of HELVETAS Swiss Intercooperation based in Nepal.

Contact: bindu.khambu@helvetas.org

The authors would like to acknowledge the continuous financial support of SDC since the 1970s for the development of the trail bridge sector in Nepal.

## REMEMBERING FORGOTTEN CROPS – DEVELOPING NEW VALUE CHAINS

The potential of so-called underutilised crops for human nutrition and as a source of income for poor farmers in the Global South was discussed at the Elsevier International Conference on Global Food Security in Cape Town, South Africa. Researchers from Kenya, Germany and the USA presented their latest research results.

By Caroline Moraza, Silke Stöber, Zoltan Ferenczi and Wolfgang Bokelmann

World-wide, the challenge is emerging to provide a growing population with the right amount of food at the right time while making sure that diets are nutritious and of high quality. Large shares of the population in developing economies continue to be confronted with insecure food supplies and, on top of this, malnutrition persists. Moreover, growing economies and low-to-middle income countries (LMIC) are experiencing an influx of non-communicable diseases (NCDs) such as diabetes, obesity and cardio-vascular disease. Therefore, agriculture and biodiversity's role in fighting malnutrition, food insecurity and growing levels of NCD incidence has gained a key status in the Sustainable Development Goals (SDGs). Next to these challenges, the SDGs also highlight the importance of biodiversity for humans. This is where the concept of agrobiodiversity – practising agriculture while maintaining biodiversity – comes in. Despite a plethora of edible plants that are known, human diets are relying more and more on a limited array of crop species to satisfy nutritional needs, resulting in a loss of diversity on fields – and on plates. This leads us to the proposition to realise the chances of reincorporating “forgotten or orphaned”, so-called underutilised crops in people's menus and on farmers' fields.

### BIG POTENTIAL, HUGE CHALLENGES

At a symposium held during the Elsevier International Conference on Global Food Security in Cape Town, South Africa, in December 2018, the comparative advantages of orphan crop value chain development (VCD) as a strategic means towards improved food and nutrition security were discussed. The symposium concentrated on what VCD interventions might be the “right” or necessary ones to foster a sustainable uptake of underutilised crops and what concrete actions or solutions could help bring back orphaned crops to people's plates. Researchers from Kenya, Germany and the USA presented their latest research results on African Indigenous Vegetables



At a local market in Nakuru, Kenya, an enumerator collects data from African indigenous vegetable (AIV) sellers for the survey panel.

Photo: Silke Stöber

(AIVs) – and, more specifically in this context, on leafy green vegetables in Kenya (see Box on page 41).

Underutilised crops are above all seen as important for food and nutrition security because they enable a diversification of food and nutrient sources, i.e. a holistic approach to human diets. In the past, as Michael Krawinkel, Professor at Justus Liebig University Giessen, Germany, explains, food and nutrition security strategies targeted calorie intake, and potential “nutrient gaps” were closed with supplementary feeding or fortification. Today, diversified diets, especially those containing fruits and vegetables on a regular basis, are widely viewed as a better way to close this gap (also see Figure on page 41).

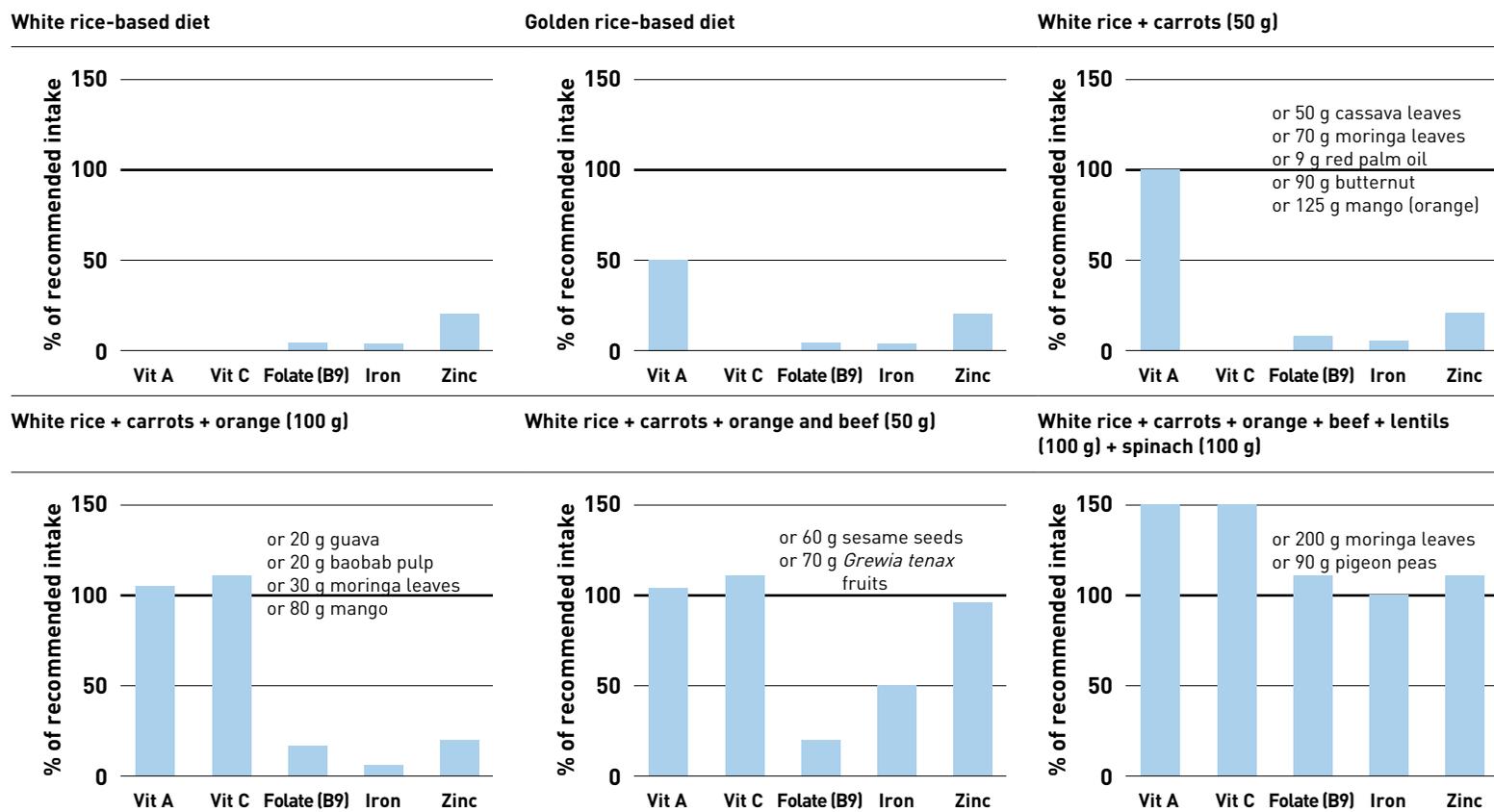
### What is happening to our diets?

To harness this potential, it is essential to build consumer awareness and sensitise people towards taking up fruit and vegetable-rich diets. The empowerment of consumers to make informed decisions to become food-sovereign stands starkly in contrast to growing consumer dependency on media and advertisements promoting the consumption of unsustainably produced and unhealthy food.

According to Elizabeth Mitcham from University of California, Davis, USA, more than half of the global population do not consume adequate amounts of fruit and vegetables on a daily basis. People tend to rely heavily on a small array of crops compared to the large va-



Step-by-step, or food item per food item, the graphs illustrate that a diet with just five different food items can already satisfy the daily requirements of five very important micronutrients.



Source: Kehlenbeck, K. (2015); data derived from WHO/FAO 2004: Vitamin & mineral requirements in human nutrition; Tang et al. (2009): Golden Rice is an effective source of vitamin A. American Journal of Clinical Nutrition 89(6).

riety of edible plants. Studies suggest a strong dietary diversity development linkage, showing a positive relationship between the availability of food groups and nutritional, health and development outcomes. Nonetheless, one growing problem spread particularly in urban centres of low and middle income countries is the dramatic increase in the consumption of processed foods as urban dwellers become more affluent. For rural areas, studies in Tanzania demonstrated that an increase in small-

holder horticultural crop production and behaviour change communication can generally increase the consumption of AIVs.

### An entrepreneurial gambit for farmers?

Integrating resource-poor smallholder farmers into growing orphan crops like AIVs holds the potential of addressing income generation in

rural areas and nutrition challenges in urban areas. Willis Kosura, a professor at the University of Nairobi, Kenya, examined the positive link between the demand for AIVs and the level of knowledge of their nutritional and health benefits. Yet, local AIV supply fails to keep pace with the growing demand. Barriers for farmers to adopt AIV production include poor infrastructure, lack of government involvement, and prevailing negative consumer perceptions of AIVs as “poor man’s food”. Kosura’s field research in Western Kenya suggests that awareness creation for the market potential of AIVs, improving information access and collective bargaining through farmer groups have the power to develop both the production and marketing ends of the AIV value chain.

### Considering all dimensions of sustainability

When considering the impacts of climate change and ecological sustainability, there are many entry points along the value chain of AIVs to turn it into a more sustainable one. Starting at the production level, the concept of sustainable intensification of farming pro-

## LEAFY GREEN VEGETABLES AND VALUE CHAIN DEVELOPMENT (VCD)

African Indigenous Vegetables (AIVs) are found across sub-Saharan Africa, but in our research we focus on leafy green vegetables commonly eaten in East African countries such as amaranth and cowpea leaves (*Vigna unguiculata*), African nightshade (*Solanum villosum*), spider plant (*Cleome gynandra*), Ethiopian kale (*Brassica carinata*), jute mallow (*Corchorus olitorius*), slenderleaf (*Crotolaria spp.*), etc. These crops have been gaining attention in sub-Saharan Africa as an alternative crop to “exotic” or introduced horticultural crops such as collard greens, green beans or tomatoes. AIVs have the potential to improve diets, diversify farm production and generate income due to a growing yet unsatisfied local demand. Nevertheless, VCD of underutilised crops like AIVs is confronted with significant challenges such as perception problems, a lack of holistic research, poor policy frameworks and weak marketing strategies. Given such multiple challenges, only a comprehensive VCD approach is able to address the obstacles found at each stage of the value chain – as the motto “from farm to plate” suggests.

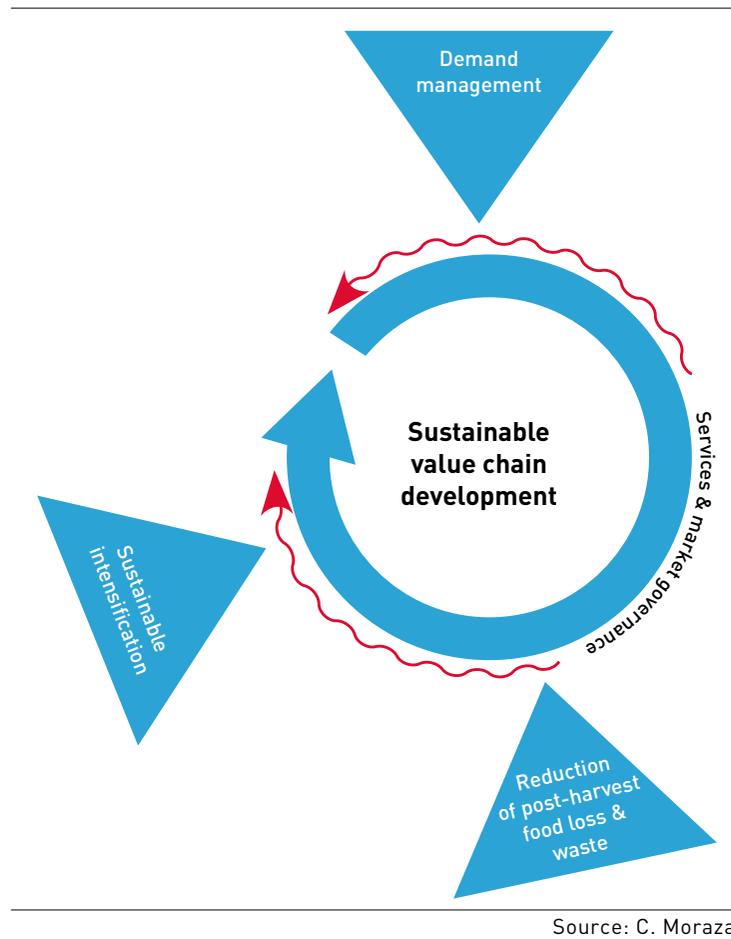
vides a general framework aiming to produce a maximum output by optimising climate and livelihood trade-offs. According to Barnabas Kurgat, a PhD student at the Centre for Rural Development (SLE) in Berlin/Germany, nitrogen fertiliser efficiency is a major hotspot in mitigating greenhouse gas emissions from AIV value chains. In his studies, a mixed fertilisation strategy with the combined use of manure and synthetic nitrogen proved to have the best economic and ecological trade-off. Still to this day, small-scale irrigation and improved seed varieties are rarely used, although they have the potential of jumpstarting the Kenyan AIV value chain. In addition, increased on-farm agrobiodiversity holds much potential for smallholder AIV producers. The climate change adaptation and ecological sustainability of AIV value chains research group at the Centre for Rural Development (SLE) has found out that AIV commercialisation need not mean a loss in agrobiodiversity. Commercial Kenyan AIV farmers in fact have the most diversified AIV species production portfolio, suggesting that specialised AIV farmers also see a potential in growing a range of AIVs.

But if there is so much potential for forgotten crop VCD to let producers make their businesses more economically and ecologically sustainable, why aren't they picking it up? Focus group discussions and household survey analysis have revealed that Kenyan AIV smallholders often have low adaptive capacities. Adaptation gaps vary, but particularly for rural areas, too little income diversification and financial management as well as limited access to high value/formal markets are major constraints.

## HOW DO WE MAKE SURE ORPHAN CROPS GET TO THE PLATE?

Although AIVs are indeed found to be climate-resilient and socially inclusive, further challenges in the mid- and down-stream segments of the value chain can hinder unlocking this potential. Challenges more on a macro-level, such as poorly organised governance of markets, land and services, increase losses of AIV in the mid-stream value chain segments. Further down-stream, at the consumer level, demand management promoting the concept

### Entry points for a sustainable AIV value chain



Source: C. Moraza

of diets rich in healthy fruit and vegetables replacing a trend towards animal protein-intense diets could help uplift the upper end and the sustainability of the AIV value chain. To make it sustainable, it is paramount to achieve sustainability along the entire AIV value chain – from consumers to traders and processors all the way to producers.

### Making post-harvest losses tangible

In comparison to cereals or tubers, leafy horticultural crops such as AIVs are highly perishable and are thus subject to heavy post-harvest losses. Preventing these losses is growing in importance not only in industrialised countries, but also in developing economies, as food losses greatly contribute to food insecurity, poverty, slow economic growth and climate change. Research experiments conducted in Kenya by Elisha Gogo of Humboldt University Berlin, Germany, have shown that losses between harvesting up until marketing can even rise to 50 per cent. Not only is a loss of quantity detrimental to upward mobility, but loss of nutritional food quality is also a huge problem for society as a whole. Poor handling along the value chain can indeed significantly

decrease the precious nutritive value (e.g. provitamin A, iron, zinc) of AIVs, which is beneficial for consumer health.

To confront this pitfall, appropriate post-harvest technologies can truly contribute to preventing losses. To achieve this, we need to revolutionise packaging and handling. Farmers and traders in Kenya commonly pack or even stuff large quantities of the fragile AIV into large bags. However, the full extent of the causes and magnitude of post-harvest losses is still not fully uncovered. In a stakeholder experiment, to create awareness for the impacts of post-harvest losses, it became evident that as soon as the economic losses were broken down into actual costs and made visible per value chain step (at harvest, after harvest, at market), farmers and traders started to grasp the magnitude and importance of post-harvest handling for their enterprise. Making economic impacts of post-harvest losses observable at the micro-level could even also help promote the introduction of new emerging post-harvest technologies,

such as new packaging material (modified atmosphere packaging, biodegradable packaging) or UV-C treatment at the market level.

### Let's talk about nutrition

Coming to the last stage of an orphan crop value chain, we encounter an important stakeholder – the consumer. Developing this stage of the value chain is crucial not only to influencing supply (AIV farmers) and demand (AIV consumers), but also in addressing consumer health issues. The paramount prerequisite is to ensure that there is a demand for an orphan crop like AIV, so it makes sense for farmers to endeavour in uptake of AIV production. One key to driving demand for forgotten crops like AIV is banking on their nutritional value. In Kenya, consumers appear to be very much aware of the nutritional benefits of AIV and are commonly eager to buy them for this reason. A research team led by Susanne Huyskens-Keil from Humboldt University Berlin investigated the impact of different handling and preparation methods on the nutritional value of AIVs. Surprisingly, the studies reveal that AIV sold in Kenya at open markets retained higher mineral composition as well as antioxidative compounds than those



After repackaging into smaller bags, the vegetables are ready for transport by public bus to customers in Nairobi who have pre-ordered them.

Photo: Silke Stöber

sold at supermarkets. Another striking result is that AIV processing (cooking and drying) does not necessarily have a detrimental effect on the nutritional content. For example, from her experiments, PhD candidate Grace Odon-go concludes that the AIV Ethiopian kale (*Brassica carinata*) is still able to protect against aflatoxin-induced DNA damage in liver cells after cooking. Aflatoxin poisoning through food consumption and its damage to the liver is indeed a highly relevant topic in developing and upcoming economies like Kenya. A lack of institutionalised quality control often entails food infested with the aflatoxin-causing fungi making it onto consumers' plates. Studies by Eliud Wafula from the Max Rubner Institute, Karlsruhe, Germany suggest that AIV processing through fermentation also shows potential to prevent the loss of nutritional compounds.

An orphan crop like AIV starting to enjoy a surge in popularity based on its nutritional

benefits gives us reason to consider it at least in its home region as an undiscovered so-called "super food". Further post-harvest treatment and processing innovations, e.g. deriving extracts or powder for the development of new functional food, might be additional steps to secure the nutritional benefits and create new marketing strategies for this underutilised crop.

**Caroline Moraza** and **Dr Silke Stöber** are research associates at the Centre for Rural Development (SLE) at Humboldt-Universität zu Berlin.

**Zoltan Ferenczi** is a research associate and **Wolfgang Bokelmann** is a professor at the Albrecht-Daniel-Thaer Institute of Horticultural and Agricultural Sciences at Humboldt-Universität zu Berlin.

For a list of references, see the online version of this article at: [www.rural21.com](http://www.rural21.com)

## Imprint

**Rural 21 –**  
The International Journal  
for Rural Development

**Published by:**  
DLG-Verlag GmbH  
Frankfurt, Germany

**Advisory council:**  
Ueli Mauderli, DEZA  
Dr Reinhard Grandke, DLG  
Louise Lutikholt, HELVETAS  
Dr Stefan Schmitz, BMZ  
Christel Weller-Molongua, GIZ

**Editorial committee:**  
Dr Michael Brüntrup, DIE  
Dorothea Hohengarten, GIZ  
Karl-Martin Lüth, DLG  
Prof. Dr Matin Qaim,  
University of Göttingen  
Dr Detlef Virchow, Plan International

**Editor in chief:**  
Silvia Richter, [s.richter@dlg.org](mailto:s.richter@dlg.org)

**Editorial staff:**  
Olive Bexten, [o.bexten@dlg.org](mailto:o.bexten@dlg.org)  
Daniela Böhm, [d.boehm@dlg.org](mailto:d.boehm@dlg.org)  
Ines Lechner, [i.lechner@dlg.org](mailto:i.lechner@dlg.org)  
Angelika Wilcke, [a.wilcke@dlg.org](mailto:a.wilcke@dlg.org)

**Editorial assistance:**  
Michael Gardner

**Translated by:**  
Christopher Hay  
Tazir International Services

**Cover photo:**  
Jörg Böhling

**Design & Layout:**  
Andrea Trapani, DLG-Verlag

**Editorial office, distribution, advertising:**  
DLG-Verlag GmbH  
Eschborner Landstraße 122  
60489 Frankfurt, Germany

**Printed by:**  
Brülsche Universitäts-  
druckerei GmbH & Co KG  
35396 Gießen, Germany

Rural 21 is published four times per year. The subscription rate is 33.– Euro (Germany), 37.– Euro (EU countries), 51.– Euro (non-EU countries) and 8.30 Euro per issue, plus postage. All rights reserved. The contents may not be translated, reproduced in whole or in part, nor may information be passed on to third parties without permission of the publisher. Please direct all correspondence to the editor in chief.

The opinions expressed by the authors are not necessarily those of the publisher or the editor. The editor retains the right to make editorial changes.

ClimatePartner<sup>o</sup>  
klimaneutral

Druck | ID 11599-1712-1001



# STAY INFORMED

WITH OUR FRESHLY DESIGNED NEWSLETTER



Verlag

March 2018/1

## RURAL21

The International Journal for Rural Development

### Dear Reader,

Please find below an overview of the topics that Rural 21 has covered in the past few weeks and events coming up in the next few months. We wish you an informative read and we look forward to your feedback!

Your Rural 21-Team: Olive Beiden, Daniela Böhm, Ines Lechner, Silvia Richter and Angelika Wicke.

### Chères et chers lecteurs,

Vous trouverez ci-dessous un aperçu des sujets couverts par Rural 21 ces dernières semaines. Les textes français se trouvent à la fin de la Newsletter. Nous vous souhaitons une lecture enrichissante et recevrons avec plaisir vos commentaires !

Votre équipe Rural 21 : Olive Beiden, Daniela Böhm, Ines Lechner, Silvia Richter et Angelika Wicke.

### News



#### FAO and OECD call for responsible investment in agriculture

The OECD and the FAO have launched a pilot project to kick-start the practical application of the OECD-FAO Guidance for Responsible Agricultural Supply Chains in small and medium-sized enterprises.

[www.rural21.com](http://www.rural21.com)



[www.facebook.com/rural21](https://www.facebook.com/rural21)

A FORTNIGHTLY UPDATE ON RURAL DEVELOPMENT