Monitoring and evaluation within FARMAF

FARMAF Working Paper

Kees Burger, Elodie Maître D'Hôtel

13 March 2012

The official *Description* of the project set as a goal of the kick-off planning meeting the adoption of 'validated monitoring & evaluation plans'. This working paper is a methodological note that outlines the approach towards setting up these plans.

This note distinguishes two levels of monitoring and evaluation (M&E). One is the more traditional level of the project itself, the other is the M&E (or impact assessment) of the tools that should serve the agricultural community in mitigating their exposure to risks.

1. The project as a whole

The formal type of M&E requires that the execution of the project be monitored and evaluated against the list of activities and goals proposed by the project. It answers the question if the project has done what it set out to do, and to what extent the objectives have been reached.

There is a formal side to this, and a more down-to-earth side. The formal side refers to the explicit indicators mentioned by the logistic framework of the project document . The indicators must be quantified; there is no specific requirement to 'prove' the attribution of the changes in indicators to the project as such. The attribution question comes back at the level of the tools: see below. What the project as a whole will actually achieve in the countries concerned, for example the changes brought about in the farm organisations, trading community, service providers etc., should be discussed and agreed upon by all stakeholders in each country.

The logframe of the project is guiding the selection of formal indicators. It set as goals an improvement in the livelihoods of the farmers and a 6% annual growth of the sector in the countries [and a 10% farm income growth after 10 years, i.e. beyond our project's lifetime]; more specifically, the objective is that by the end of the project, 10% of the farmers have access to the tools, that their use of yield-enhancing inputs increases by 25%, and their crop yields grow by 25% too.

For M&E, these objectives translate into the requirement that measures of these indicators are available at the start and at the end of the project. As we do not (yet) know who the users of the tools will be, the initial measure of use of inputs and crop yields, should refer to all farmers, with a possible restriction to those farmers that grow crops relevant for the project. Relevant crop of the project is 'grains', i.e. maize in all three countries, perhaps extended to rice, sorghum, and wheat. Thus, the first step will be to ascertain the most recent level of

yields and input use of the grain farmers in the three countries. If the information is not available at an adequate level of reliability or detail, a targeted survey should be considered.

In addition, a baseline estimate must be made of the extent to which these farmers presently make use of the instruments targeted by the project: crop insurance, (credit based on) warehouse receipts, centrally governed market information systems, a commodity exchange and collective marketing. Partly, this use is indirect: if traders utilize the commodity exchange, say, then their suppliers may indirectly benefit. As a baseline, then, we need an estimate of direct and indirect use. This can probably be derived from the present uptake of crop insurance, WRS-based credit, clients of commodity exchange or MIS-service, and – in case of traders and processors – their supplying farmers, complemented by a scan of the extent to which farmers use collective marketing of the targeted crops. These numbers of farmers would constitute the baseline, and the numbers of smallholders among them are targeted to amount to at least 10% of the (smallholder) farm population as per the project's objective.

Eventually, this information will need to be updated toward to end of the project. The same type of sources that supplied the information for the baseline can be used for the final assessment.

At a more instrumental level, these objectives of the project are to be reached – according to the logframe – through national and regional action plans, collaboration arrangements; improved capacity of NFOs and service providers; better systems for WRS, MIS and commodity exchange; improved insurance products and material that synthesizes the successful approaches. To substantiate these expected results of the project, an inventory must soon be made of the present state of affairs in the countries. These inventories should elaborate on

- the present organisation of the provision of the services to the farm sector (activities of government, NFOs, industry; coordination mechanisms among them)
- the present capacity of these entities to deal with the targeted tools (using as indicators the numbers of people with expertise in insurance products, MIS, WRS etc. in these organisations; or the turn over; or proxy indicators for the same; or the shares passing through commodity exchanges and so on)
- present types of instruments available: description of insurance policies that are offered, regulations and modalities of commodity exchanges, WRS, MIS; and measures for their use and adequacy
- review of the written information on the presently accessible tools

This information can be obtained through contacts with the organisations and service providers involved. To contribute to the M&E of the project, the information should be updated on an annual basis. After four years, it should show the changes made in this period and link to the objectives considered above.

At the level of the project, it seems that M&E does not require a great deal of work beyond what can be obtained from the participating organizations. Only if this information is inadequate, targeted surveys among farmers and traders are in order.

As to responsibilities, these are probably best put on the plate of the coordinator, who could delegate some of this to the country coordinators, the NFOs (national manager) and/or to the thematic 'leaders'.

Annual targets and milestones for the indicators should be elaborated in the national action plans. For the eventual indicators, that are needed by the end of the project, annual reporting is not necessary but can be helpful; at the instrumental level, however, annual M&E is required: what stage of improvement, development or implementation do the national coordinators and other responsible persons want the project to be in in any of the next four years? These milestones should be formulated soon after the base line inventory is made.

2. The level of the tools

The other level at which M&E is envisaged is that of the tools themselves, and this requires some more consideration.

The program addresses five types of risk management tool: insurance at the farm level, warehouse receipt schemes at the trader's level and, wherever applicable, improved Market Information Systems, commodity exchanges in addition to collective marketing by farm groups. For each of these tools, the project aims at some type of improvement, in some cases through product development, in other cases through more expert promotion, of better regulation or other means. Point is, the project expects to accomplish a change, and it is not yet clear how best to bring this about.

M&E should be employed to find out what would be the 'best' way. For this purpose the approach towards 'tentatively' promoting each type of tool should be set up in such a way that a justified conclusion can be drawn from the activities and measurable outcomes.

The *Description* requires this from the project: under 'Expected result 5' and in WP5, evidence based knowledge is said to be generated by the project concerning the tools and the context in which these can be successful. To generate such evidence, a dedicated methodology is required that enables attribution of outcomes to deployment of tools, and that also sheds light on the causes for any success of such tools.

Crucial element in the methodology is the generation of data on the counterfactual: what would have happened in the target group had the tool not been introduced? Answering this question requires – next to a precise definition of the tool and of the target group – a convincing story about how well the unobservable outcomes can be approximated by observable outcomes. Note also that the question is not what households would have done had they not adopted the tool. This latter question is only part of the answer. Success of a tool is measured by rate of adoption, and the outcomes of the actions subsequently undertaken by those that adopt the tool.

Both elements, adoption and outcomes of actions of adopters, are measured in a context that changes from day to day and construction of a good counterfactual should take this into account. A mere comparison of before-after may suffer from changes in these circumstances (prices, policies, weather, etc.). A comparison with a group that is not exposed to the tool is a better option, but even here comparison should be made in terms of differences over time: did the group exposed to the tool change in ways different from the group not exposed to the tool. This so called dif-in-dif approach requires collection of data in the group before and after exposure to the tool, *and* in a comparable group that is not in a position to use the tool.

Thus, to assess a tool and thereby generate evidence on its suitability for up-scaling, four steps plus one must be taken

1. identify the tool. This seems obvious, but can be tricky: in the case of weather-index based insurance, the tool could be the index, not the insurance policy; or it could be an index-insurance-linked credit line that is opened; or indeed the insurance policy as such. In the case of WRS, the tool can be the facility itself, or the governance structure, or the grading system.

2. identify the target population or initial sub-population for testing the tool. The initial population should be representative of the eventual population.

3. establish the way impact is to be measured: what indicators? This implies to build a 'theory of change' and derive potential indicators

4. Collect data on those indicators before the introduction of the tool in both 'treatment and 'control' groups. These tool-specific baseline surveys are bound to be in October-

November 2012. Data collected during the baseline surveys are different from data collected during the scoping missions (micro-economic data).

At this point, the experiment can start: the tool is made available in one group but not (yet) to the other. The rate of adoption and the impact on the households (indicators) will tell to what extent the tool is successful. Time horizon is important: it may take some time before knowledge on the tool is sufficiently wide spread and it can be useful to allow for one or two years to check adoption rates:

5. Collect data on those indicators at time t+3 (end of 2015): who are the adopters, what are their characteristics, and how did the adoption change their livelihoods. This information is derived from a comparison of adopters and non-adopters in the treatment sample with similar households (matching) in the control sample. This information on (non-)adopting households is also needed to prepare for up-scaling of the tool to wider audiences than the population for which the initial test-population is representative; it helps in documenting the deeper effects of adopting the tools so that future users can be adequately informed.

activities	period	Description
Task 1	August 2012	identify the tools
Task 2.	September 2012	identify 'treatment' and 'control' groups and the samples to measure adoption and the impact
Task 3.	October 2012	Agree on indicators and their measures for impact assessment (indicators)
Task 4	November 2012	Baseline survey in both groups
Task 5	November 2015	Follow up survey in both groups

If tools are already in the field and working and the project seeks to improve the rate of adoption or improve the tools itself, the above set-up can be adjusted. It will not in all cases be possible to define a good 'control' group anymore, because the initial adoption and marketing may have targeted groups of households with specific characteristics (say larger farmers). In these cases a second-best approach is to be followed that closest resembles the treatment-control design.

The above approach is suitable for the assessment of tools that are acquired by farm households at a price. Only then does adoption stand for a certain expected usefulness in the eyes of the household. If tools are free of charge and require little effort, their rates of adoption may not be a good measure for their usefulness, for the adoption may not lead to any change for the household.

This is typically the case with public goods. Households can use these often free of charge. Making these goods available entails (social) costs, but adoption rates cannot tell if the benefits outweigh the costs. In these cases, the information as collected in step 4 provides the answer: the impact that adoption had on the households (measured typically by a dif-in-dif approach using similar households in the control group) can tell whether the public good or the freely (cheaply) provided tool was useful and efficient.

Finally, in cases where a treatment and control group division cannot be established, how can one generate evidence on success of tools? An example is the launch of a Market Information System, which is difficult to make exclusively available for one group. Generating

evidence in this case rests on third-best alternatives, such as comparison with countries that do not have the system yet, comparison with earlier years, and comparison with other crops for which no MIS is active. In all cases, the comparison is with scores of agents that find themselves in circumstances (beyond their control) without MIS. The comparison is not between individuals that use an existing MIS and those that do not. Such comparison would suffer from serious selection bias.

In all cases, whether the tool is a public good or a private, excludable good, there is a possibility to assess its effects through games. If appropriate groups are selected to participate in games that ask the players to simulate real life trading, useful lessons might be learned about optimal design of the trading system or MIS. This also generates the documented evidence asked for by the project's description.

3. Some considerations for the various tools

Below follows some elaboration for the types of tool that are included in the project.

Insurance products

In this case, where product development is an important element, the ideal approach would be to differentiate the product and/or the its marketing, and put it into a randomized controlled trial. This will not be possible in many cases, but its core element should be maintained: that eventual differences between groups can be attributed to the intervention. For example, an insurance policy can define its fee and pay-out in two ways A and B, say one in kind and one in cash. The project can help to find out which of the two policies has a better uptake by the farmers (as a first measure of success), by finding a good setting where some farmers are exposed to product A and some to product B, and some function as control group. In close collaboration with insurance companies and the NFOs a range of products (differentiated into various modalities for fees/conditions/timing, marketing etc.) can be developed for testing in the field through randomized controlled trials, or similar approaches.

To learn from these experiments, repeated farm surveys are needed. A target population should be set from the start. Alternatively, one can measure potential uptake from sampled responses after informational meetings, but this would be less reliable. As Giel Ton[1] suggests, this quantitative information should be complemented with case studies and be preceded by a common understanding on the relevance of the chosen indicators for subsequent selection of the preferred tool and its up-scaling.

The same approach can be followed for inputs into the product's definition: a weather index based insurance can use fine or wide grids for the rainfall indicator, with implications for the correlation with farmers' incomes and the probabilities of receiving payments, and – from this – for the fees to be paid. An experimental approach – either a field experiment or the same in a game setting - can then be used to find out which insurance would be more appealing to the farmers, which group of farmers would be interested in either one. The latter information helps in assessing any selection bias that may occur.

Similarly, the industry and the NFO may find it difficult to know how best to combine insurance with credit. An experimental approach may again help in answering this question and from there, make a better start with the broad introduction of a product.

Warehouse receipt schemes, commodity exchanges, collective marketing

---

[1] Ton, Giel (2012) The mixing of methods: A three-step process for improving rigour in impact evaluations. *Evaluation*, 18(1) pp. 5-25.

In this case, the situation is somewhat different as (arguably) the target population is not just farmers, but also traders. In some cases where a substantial share of credit to farmers comes from traders, better access of (primary) traders to credit may help farmers too. Assessing the best way of setting up the WRS cannot easily be done experimentally in a real life setting, because of the costs involved and the need to include the whole trading community in adopting the rules and facilities. But the merits of various ways to set up the schemes can be assessed in an game setting, involving farmers, traders etc. By selecting representative groups, and devising relevant rules for the games, useful insights might be obtained on the interest of various groups in using WRS.

Apart from this experimental approach to developing an optimal form for WRS, there is a need for M&E to assess the real life practice of using WRS as a basis for credit provision and price discovery on commodity exchanges. Comparing users and non-users would run into problems of self-selection. Comparing one commodity with another will yield differences that cannot be easily attributed to the WRS/Commodity exchange per se. Some possibilities are offered by a before-after comparison, for which baseline data should be collected before a major change in the facilities would apply, and ex post data would help assess the effects. A problem of attribution will remain, but comparison with another commodity or another country (for which facilities did not change) may help. This amounts to a difference-in-difference approach, and requires data collection on the commodity of study *and* on the 'control' commodity or the 'control' country.

Data required are data on indicators of the degree of success of the scheme. Agreement on these indicators must be established ex ante, of course. These can be share of farmers/traders using WRS, share of credit based on WRS, efficiency of WRS per se, share of trade going through the exchange, measures of confidence that stakeholders have in the scheme etc.

Similar arguments apply to the issue of collective marketing. Experimental tools can be used in a game setting to test what types of rule would appeal to prospective participants, while the eventual group-based selling or buying activities can be monitored and assessed by deciding ex ante on indicators, comparability of other groups or groups for other commodities. Again, a difference-in-difference approach appears a sound basis for assessment and attribution. Several case studies should complement the quantitative approach to help clarify the causal patterns and thereby the significance of the findings for use in other contexts.

4. General considerations

In all M&E approaches, agreement ex ante among the stakeholders is a precondition for successful application. As noted in Kusak and Rist[2], the first step is assessment of readiness. This readiness to embark on an evidence-based approach toward improving access to risk management tools needs to be present among all participants in the project, within AGRINATURA and between it and its partners, and with the other partners in the three countries. The launch workshop would be a good place to agree on this. The next steps (of the ten) are to agree on outcomes, on indicators for these outcomes, on how to measure baselines for these indicators, and what targets to achieve. The national planning workshops could usefully have these items on the agendas.

---

[2] Jody Zall Kusek and Ray C. Rist (2004) Ten steps to a results-based monitoring and evaluation system : a handbook for development practitioners . World Bank.