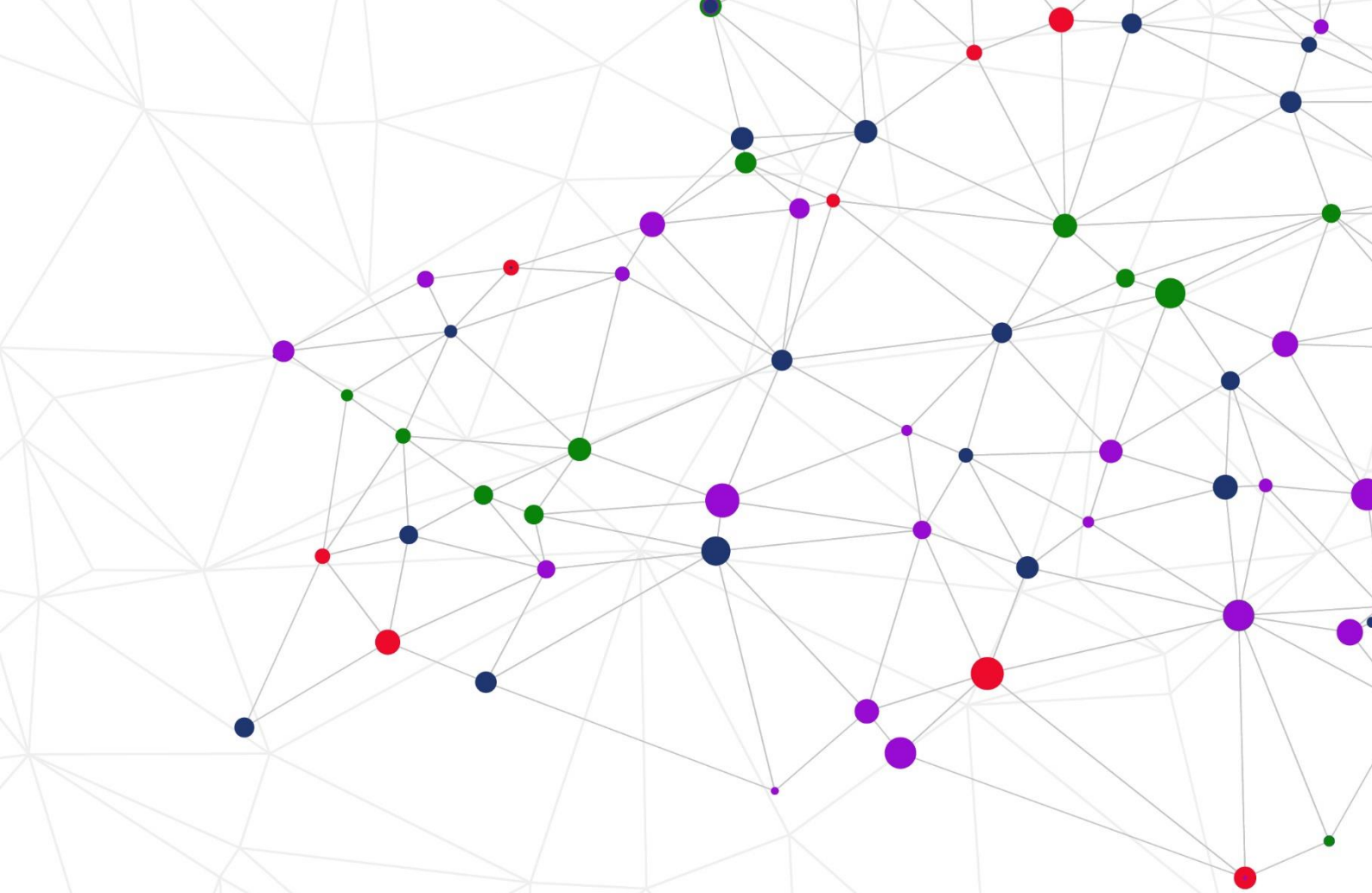

IFAD's Methods for **Impact Assessments**

A summary note

IFAD
Investing in rural people

RESEARCH AND IMPACT

IFAD's Methods for

Impact Assessments

A summary note[1]

1 This note is produced by Aslihan Arslan and Romina Cavatassi in the Research and Impact Assessment (RIA) Division of the Strategy and knowledge Department of IFAD. For more information: a.arslan@ifad.org; riamailbox@ifad.org

I. Background and objective

IFAD measures impacts of its investments by systematically conducting impact assessments (IAs) on a sample of about 15 per cent of projects selected among those that close during each replenishment. IFAD's corporate IA methodology was pioneered under the IFAD9 Impact Assessment (IA) Initiative; systematized, standardized and improved in IFAD10 (the tenth replenishment of the IFAD); and further strengthened in IFAD11.

The overall objective of this note is to summarize the methodology used for these IAs as well as how they are used to measure impact for IFAD.

II. Data collection and analysis methods

The methodology relies on ex-post quasi-experimental IAs and allows the estimation of projects' impacts that are aggregated through meta-analysis and extrapolated to the entire portfolio of projects closing during each replenishment period. This allows IFAD to estimate corporate level achievements against Tier II development indicators' targets.[2] State of the art sensitivity analyses and robustness checks ensure that the findings are reliable.

a) Selection of projects for IAs:

IFAD's sample selection requires a protocol to select projects to undergo an IA while meeting a set of criteria. The sample selection protocol used for IFAD11 IA is laid out in what follows:

The first step of the protocol was conducted in July 2018. It requires selecting at least 15% of projects from the full list of those expected to close during each replenishment period (which is called the "IA universe").

Once the list is drawn **the second step** of the selection protocol requires that the following set of inclusion and exclusion criteria are used to ensure feasibility and rigor:

i. **potential to learn lessons;**

Given that the IAs are *ex-post,* the project designs take place on average 8-10 years before sample selection.[3] In light of the importance to conduct IAs that can feed lessons learned into new project design, *ceteris paribus* projects that provide an opportunity to learn from thematic, design or implementation points of view are preferred during sample selection.

ii. **feasibility of conducting a scientifically rigorous IA;**

In some cases it is very difficult to identify a suitable comparison group, for example when the project promoted technologies, practices or financial services nationwide making it challenging if not impossible to identify areas or households not exposed to project. This has happened for example with a project that had promoted climate smart agriculture in a small country over two phases. Project areas were spread across the entire country with large spillover effects, leading to difficulties in identifying suitable comparable villages and/or households. Such projects are excluded from the sample.

iii. **buy-in from the government and IFAD;**

Access to project M&E data, targeting criteria and approaches used in selecting beneficiaries is crucial for an IA to be properly done. In cases where there is no government buy-in, or when support from PMU cannot be guaranteed it is very difficult to conduct an IA. Therefore projects with buy-in are preferred.

iv. **the capacity of a project to represent IFAD's overall portfolio**

It is important that the selected projects are as similar as possible to most of the projects IFAD supports. As such, very specific projects that are rare or unique in their logic, components or approaches are not replicable or extendable to the rest of the portfolio. Such projects are excluded from the sample.

[2] IFAD12 Results Measurement Framework (RMF) Document (IFAD12/3/R.2/Add.1). See also RMF dashboard: https://www.ifad.org/en/rmf-dashboard

[3] In IFAD11 IA, the average length of projects was 8.5 years and project design process takes around 2 years.

v. **the relevance of the IA for subsequent project phases.**

Given the goal of feeding lessons learned from IAs into future design and implementation, projects that are possibly going to be replicated either in the same country or in other countries with a similar approach are preferred, *ceteris paribus*, for the IA sample.

In the **third step,** projects meeting these criteria are listed by region and discussed with regional divisions to confirm the inclusion and exclusion criteria. Exclusion criteria are attributes that disqualify projects from being included in the sample. These include – for IFAD11 – attributes such as: a) the closing date has been postponed beyond 2021, but not yet officially reported; b) local or national conflicts were expected to prevent data collection; c) there was no government buy-in to undertake the data collection (i.e. failure to meet the eligibility conditions); and d) project was expanding to new geographic areas and less than 70 per cent of the budget had been disbursed at the time of selection. Any other reason, such as "project not performing well," has been explicitly ruled out as an exclusion criteria and all such projects have been included.

In the **fourth step** the final list of projects is prepared. The number of selected projects per region must be proportional to regional share of projects in the universe, and the remaining projects make up the replacement list to be used if any of the selected projects eventually drops out (for reasons similar to the exclusion criteria reported in step three).

Starting in IFAD12 the sample selection for the IA is done using random sampling stratified by region following the recommendations of IFAD's Executive Board. The exclusion criteria above are still applied and a replacement project is selected from the replacement list.

b) Household sample and data collection:

The main challenge for identifying impacts is to find a valid comparison group that is as similar as possible to the treatment group in terms of pre-project characteristics (and hence can represent the state of nature for the treated in the absence of the program under certain assumptions, i.e. counterfactual). When the only difference between the treatment and comparison groups is that the members of the treatment group receive the project's activities, while the members of the comparison group do not, the observed difference in outcomes can be entirely attributed to the program and the causal impact can be identified (Ravallion, 2005).

To achieve this objective, once the projects in the IA sample are selected, IAs are implemented using quasi-experimental methodologies.[4] This approach takes into account the following, inter alia: careful development of IA plans in collaboration with project stakeholders; extensive fieldwork for tailoring, contextualizing, finalizing, and piloting the field survey questionnaire; constructing a sampling frame including treatment and comparison sites and households.

The sampling design is one of the most important steps to ensure a robust counterfactual and is implemented using statistical matching techniques and validation procedures at multiple levels including geo-referenced data combined with socio-economic data. The first level requires identifying comparison sites to the beneficiaries' sites listing eligible locations given project's targeting criteria. Once these are identified data from GIS sources (climatic, agro-ecological, market and population data), available censuses (agricultural or population) as well as other available data are used to select comparable sites through propensity score matching. These are validated through expert consultations and the final list of sites is randomly selected for beneficiaries and comparison groups. A replacement list is also selected randomly in case any site cannot be reached for any reason (e.g. including natural hazards, social conflict).

Once the sampled sites are identified, households to be interviewed are selected randomly from the full list of households generated both in treated as well as in comparison sites.5 Household surveys are then conducted through Computer Assisted Personal Interviewing (CAPI) tools using tablets to collect quantitative data.6

[4] Quasi-experimental IAs are defined as those for which treatment is not randomized and a robust counterfactual/comparison group (that is as similar as possible to the treatment group in terms of pre-intervention characteristics) is created using statistically robust methodologies to identify causal impact (Angrist and Pischke 2010; White and Sabarwal 2014).

[5] The number of households to be interviewed is calculated using sample size calculations to detect the project's impacts as specified in Log Frames.

[6] Anonymized household data are currently stored in the FAO microdata catalogue: IFAD Impact Assessment Surveys (fao.org)

Qualitative data coming from key informant interviews and focus group discussions are also collected to fine tune the questionnaire, validate and better understand findings.

c) Impact analysis on a sample to proxy the population:

Project level IAs are conducted using robust impact estimation methodologies to estimate project impacts on a large set of indicators measuring IFAD's goal, strategic objectives and mainstreaming goals, as well as impact channels to maximize learning.[7] Multiple estimation specifications are used to establish robustness.[8]

The IAs are conducted on a sample selected from the full list of projects closing in each replenishment period (i.e. universe), which is used to proxy the population. Statistical analyses are conducted to establish whether the sampled projects are statistically significantly different from the universe in terms of project ratings and characteristics. Once it is established that the sample is not different from the universe, the potential for *ex-ante* selection bias can be ruled out. If this cannot be established, statistical methodologies to correct for sample selection are employed to assess the robustness and sensitivity of results.

d) Aggregation:

IFAD's aggregate development effectiveness is measured using a meta-analysis of individual project impact estimates.[9] Meta-analysis outcomes are treatment effects (mean effect sizes) and represent the impact of projects in the IA sample. They represent attributable impacts for the sample and can be also interpreted as percentage changes over comparison groups for each development impact indicator. Meta-analyses results are used to compute aggregate corporate impacts. For IFAD11, the IA sample is not systematically different from the universe of projects from which it is selected, therefore the mean effect size from the meta-analysis is extended to the entire IFAD investments that closed during the replenishment period.

e) Sensitivity analysis:

The mean effect sizes from the meta-analyses are validated by estimating impacts using the *pooled* household level data. IFAD re-runs analyses by combining all individual IA micro-level data together and running a pooled data analysis, which controls for country/project level unobserved characteristics that may influence impacts.[10] The data, programs, and other details of the computations sufficient to permit replication along with their revisions are encrypted and anonymized.[11]

III. Projection methods

IFAD linked its targets to Strategic Objectives and Economic Goal with the Development Effectiveness Framework. The results of the meta-analysis are used to calculate these targets: number of beneficiaries with increased productive capacities (SO1), increased market access (SO2), stronger resilience (SO3) and better nutrition (starting in IFAD11).

In IFAD10 IA and IFAD11 IA, IFAD projected the aggregate meta-effects to the universe of its closed projects. In IFAD11 IA, these were 96 projects that had a total cost of US$7.1 billion (including international and national co-financing). The projection requires determining the number of targeted beneficiaries across the universe of

[7] It is important to note that IFAD collects data from beneficiaries that were targeted by the entire project/investment, i.e. our data collection does not distinguish between IFAD investments and co-financing.

[8] These include the doubly-robust Inverse Probability Weighted Regression Adjustment (IPWRA) method and Nearest Neighbour Matching (NNM) method, among others.

[9] Meta-analysis is a statistical procedure for combining data from multiple studies, or in the specific case of IFAD, project impact estimates. Meta-analysis can be defined as a synthesis of results or "the statistical analysis of a large collection of results for the purpose of integrating the findings" (Glass 1976). In other words, it is "a quantitative summary of statistical indicators reported in similar empirical studies" (Brander et al. 2006). See: Glass, G. (1976). Primary, Secondary, and Meta-Analysis of Research. Educational Researcher, 5(10), 3-8, and Brander L.M., et al. (2007). The recreational value of coral reefs: A meta-analysis, Ecological Economics, Vol. 63, Issue 1, 2007, 209-218.

[10] These analyses exploit the between-project variability and control for country/project specific unobserved characteristics, thus improve the external validity of the overall impact assessment. See for example Nijman. T. and M., Verbeek. 1992. "Nonresponse in Panel Data: The Impact on Estimates of a Life Cycle Consumption Function." Journal of Applied Econometrics, 7(3), 243-57.

[11] Data are available upon request which should include an explanation of how they will be used.

eligible investments, which was equal to 112 million beneficiaries for all projects in the IFAD11 IA universe.[12] The total number of beneficiaries impacted for *each outcome* is then calculated based on this number.

The projection relies on an important assumption concerning the distribution of impacts resulting from the meta-analysis. It is assumed that estimated impacts are normally distributed across the entire population of beneficiaries and have the same means and standard deviations as the estimated impacts. Since the population is large, this is not a constraining assumption. Given the distribution, the calculation of the number of people that have benefited from IFAD interventions during this period is determined by the number of beneficiaries in the distribution that exceed the threshold set in the RMF for each of the Tier II development indicators.

In summary, total number of beneficiaries who have achieved results above the target set in RMF is obtained by: 1) randomly drawing a normal distribution of impacts (with an associated mean and standard deviation as empirically estimated from the meta-analysis) for 112 million people; and 2) counting the number of people that have experienced an increase that exceeds the threshold set for the corresponding outcome (or IFAD RMF indicator).

IV. IFAD11 IA results

Results from IFAD11 IA reported in the table below, show that for each target, except for nutrition, results have surpassed the objectives set by far.

Regarding IFAD's overarching goal, investments collectively improved the incomes of 77.4 million beneficiaries by at least 10 per cent, against the total target of 44 million[13] (over three years). Productive capacities (SO1) of 62.4 million beneficiaries were improved against the target of 47 million, and the market access (SO2) of 64.4 million beneficiaries increased against a target of 46 million – in both cases, by at least 20 per cent. Around 38 million beneficiaries have seen their resilience (SO3) improve by at least 20 per cent.[14] The target of 12 million people with improved dietary diversity (of 10 per cent or more) is the only target not met during IFAD11.

Table 1. IFAD11 RMF targets and IFAD11 IA results: Number of beneficiaries above RMF target

GOAL/SO	**RMF INDICATOR**	**DEFINITION** (Threshold)	**IFAD11 TARGET** (Million People)	**IFAD11 IMPACTS** (Million People)
GOAL	2.1.1	Number of people with increased income (at least by 10%)	44	77
SO1	2.1.2	Number of people with improved production (at least by 20%)	47	62
SO2	2.1.3	Number of people with improved market access (at least by 20%)	46	64
SO3	2.1.4	Number of people with greater resilience (at least by 20%)	24	38
MAINSTREAMING GOAL	2.1.5	Number of people with improved nutrition (at least by 10%)	12	1

Source: IFAD12 RMF Document (IFAD12/3/R.2/Add.1), October 2020.

[12] Information extracted from IFAD's Operational Results Management System (ORMS).

[13] It is important to note that results from the Impact Assessment are also used to set future target and ambitions for IFAD and to assess progress towards them, including the target of achieving 40 million people with higher income per year by 2030.

[14] When a 10 per cent threshold is used, 64 million people had higher resilience compared to the control group after exposure to shocks.

V. Moving Forward

Moving forward, IFAD is committed to strengthen the methodological approach to measure corporate impact. For IFAD12 IA[15] the sample of projects used to conduct IAs has already been selected randomly and stratified by region.

To ensure that results from IA feed into new project design and strategies, lessons learned have been distilled from each IA and are being inserted into the ORMS system for project design teams to use.

In addition, a parallel effort is being made to support the data collection for the Core Outcome Indicators (COI) that are required for all projects designed from 2020 onwards. The RIA IA and COI questionnaires are being aligned to cover both key indicators needed to measure achievements against the RMF targets and those needed for project log-frames, which is an important step to ensure quality and quantity of data availability in the future.

The need for better linkages between project M&E and corporate results reporting was one of the core priorities outlined in DEF 2016. As stated in the RMF, starting in IFAD12 all projects will carry out COI surveys at baseline, midterm and completion stages. This is the best way to capture the results of a project's interventions over the course of its implementation. By linking and synchronizing M&E and IA activities as part of the survey implementation, it will be possible to conduct the IAs using a larger set of projects. Currently only COI endlines include a control group and can be used for IAs, and if COI data can cover a control group starting from baseline more robust IA methodologies would also be facilitated.[16]

This will allow a move towards measuring the Tier II development impact indicators using data from the M&E system, thereby transforming monitoring and evaluation into monitoring for evaluation (M4E). As the average duration of an IFAD project is around eight years, ideally in 2030 at the eve of the SDGs deadline, IFAD will be the only IFI to rigorously measure progress towards SDGs using its own M&E data comparing beneficiaries to a counterfactual group. This is possible only because IFAD has carefully invested in robust IAs starting in IFAD9.

[15] IFAD12 IA is defined as the Impact Assessment conducted on projects completing between 2022 and 2024.

[16] Having two rounds of data from same households would facilitate controlling unobserved household heterogeneity using difference in difference methodology to estimate impacts, which are currently based on identification using observable characteristics.

Investing in rural people

International Fund for Agricultural Development
Via Paolo di Dono, 44 – 00142 Rome, Italy
Tel: +39 06 54591 – Fax: +39 06 5043463
Email: ifad@ifad.org
http://www.ifad.org