IMPACT ASSESSMENT MANUAL

IFAD
Investing in rural people

FOREWORD

This report was produced by staff from the Research and Impact Assessment (RIA) Division of the Strategy and Knowledge Department of the International Fund for Agriculture Development (IFAD) under the guidance of Sara Savastano, Director of RIA.

This manual is designed to provide detailed step-by-step information, examples and clear guidelines to those interested in conducting rigorous impact assessments using the approach followed by the RIA division of IFAD. It aims to give the reader all the instruments for assessing the impact of a given development project, and at the same time, ensure comparability across projects and countries along the entire impact assessment (IA) cycle. A standardized set of activities and procedures is crucial for research transparency, data sharing, and aggregating estimated impacts from individual projects to calculate corporate impact.

With reference to each step of the IA cycle, the reader is given all the information as well as relevant examples and template documents to be used and adapted to project context within each phase of the assessment. The sections in this manual comprise a description of all the activities to be performed, together with a series of Relevant Readings and Key Definitions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF BOXES

ABBREVIATIONS AND ACRONYMS

TERM	MEANING
ADM	Administrative Services Division
CD	Country Director
CPM	Country Programme Manager
CPO	Country Programme Officer
ECG	Environment, Climate, Gender, and Social Inclusion
GT	Generic Tender
IA	Impact Assessment
IFAD11	Eleventh Replenishment of IFAD's Resources
LTA	Long Term Agreement
M&E	Monitoring and Evaluation
PMU	Programme/Project Management Unit
PMI	Sustainable Production, Market, and Institutions
RIA	Research and Impact Assessment Division
RIGA	Rural Income Generating Activities
SKD	Strategy and Knowledge Department
TOC	Theory of Change
UNGM	United Nations Global Marketplace

1. INTRODUCTION

The Research and Impact Assessment Division (RIA) is one of the three divisions in the Strategy and Knowledge Department (SKD) of the International Fund for Agricultural Development (IFAD). RIA leads IFAD's impact assessments with the aim of measuring the attributable impact of IFAD's investments on its beneficiaries. The evidence and insights generated inform IFAD's priorities for investments and policy engagement in order to effectively promote inclusive and sustainable rural transformation. RIA focuses particularly on research, impact assessment and policy engagement. It provides analytical inputs on internal processes to achieve development objectives (doing the right things) and design individual projects that are geared towards meeting intended objectives (doing things right). In addition to the core function of impact assessments, the division has two other functions: i) management of several research grants implemented by external partners and institutions to support **Agricultural Research for Development (AR4D);** and ii) host the data use component of the 50 x 2030 Initiative, to build the capacity and motivation of decision makers to use data and to improve data production and dissemination to meet their needs. These three functions are organized in three clusters.

The impact assessment cluster

WHAT IS IMPACT ASSESSMENT (IA)?
IA is an approach to evaluate whether observed changes in outcomes among project target groups can be attributed to development projects. Simply comparing areas with and without projects or comparing indicators before and after projects often fails to account for factors that may contribute to observed changes such as economic factors, natural disasters or conflicts.

produces rigorous evidence in the form of methodological and applied research linked to IFAD financed projects that can inform country-level investments. The focus is on learning lessons on individual projects as well as feeding into aggregate corporate reporting. The project level impact assessments were designed to allow IFAD to draw lessons from specific projects, aggregate lessons, and impact at the corporate level, and subsequently project them to the overall IFAD portfolio. Corporate reporting includes the key indicators, which are linked to IFAD's corporate goal and strategic objectives: economic mobility (corporate goal), improved productive capacity (Strategic Objective 1), improved market access (Strategic Objective 2), and greater resilience (Strategic Objective 3). These indicators complement project specific impact indicators considered as part of the individual impact assessments, which are identified in consultation with governments and IFAD regional divisions. Knowledge generated from these efforts promote effective rural development.

This manual is designed to provide detailed step-by-step information to those interested in conducting rigorous impact assessments using RIAs approach described in the **Impact Assessment (IA) Cycle**, Figure 1. **The Impact**

Assessment (IA) Cycle It aims to give the reader all the instruments for assessing the impact of a given development project, and at the same time, ensure comparability across projects and countries along the entire impact assessment cycle. A standardized set of activities and procedures is crucial for research transparency, data sharing, and aggregating estimated impacts from individual projects to calculate corporate impact.

Figure 1. The Impact Assessment (IA) Cycle
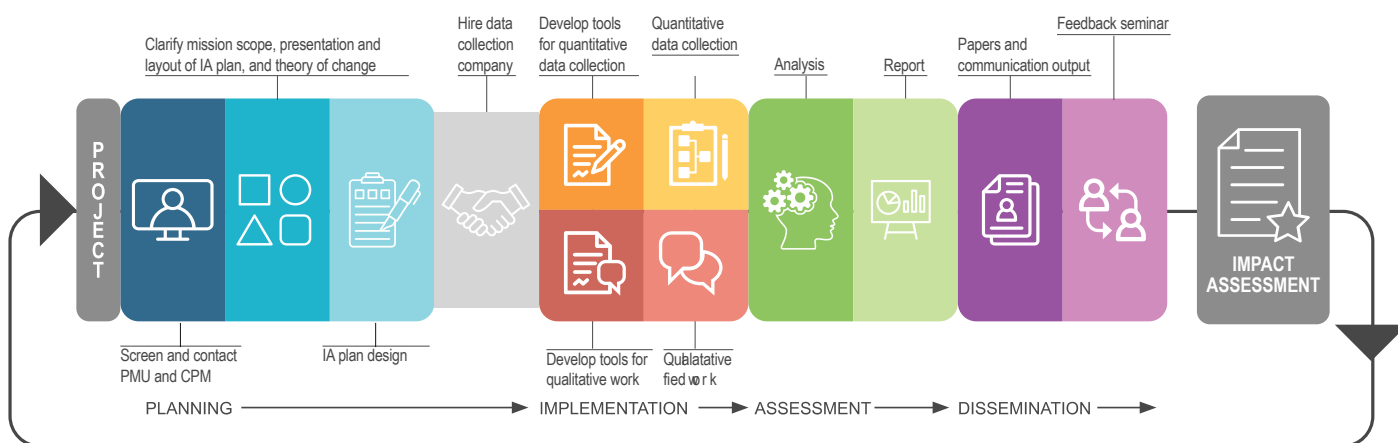
With reference to each step of the IA cycle, the reader is given all the information as well as relevant examples and template documents to be used and adapted to project context within each phase of the assessment. The sections in this manual comprise a description of all the activities to be performed.

1. IA Cycle

The IA cycle as described in Figure 1 is made of all the steps in order to evaluate the attributable impact of a project on outcomes of interest. There are four main steps, and each step includes specific activities.

KEY DEFINITION

IFAD WORKS TO BRING ABOUT DEVELOPMENT OBJECTIVES EFFECTIVELY

Effectiveness is defined as "the extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance (OECD, 2002)".

Planning: In consultation with governments and IFAD country teams, the theory of change (ToC) of the project, the indicators to be considered, and the questions to answer in any given impact assessment are clarified following the logic of the project and IFAD's main corporate indicators. This leads to the writing of a standardized Impact Assessment Plan, which identifies the quantitative and qualitative approaches to be used and the questions to be

answered in the impact assessment. Finally, a data collection firm is procured to carry out fieldwork.

Implementation: The tools for qualitative and quantitative data collection are developed, data collection protocols to ensure quality control created, the qualitative and quantitative fieldwork completed, and data sets created.

Figure 2. IA Timeline and Deliverables

ACTIVITY		MONTH											
		1	2	3	4	5	6	7	8	9	10	11	12
1	**PLANNING**												
1.1	Screening Process												
1.2	Contact with PMU and CPM	▓											
1.3	Scoping Mission		▓										
1.4	IA Plan Design			▓	▓								
1.5	Hiring Data Collection Firm				▓	▓							
2	**IMPLEMENTATION**												
2.1	Qualitative and Quantitative Tools				▓	▓							
2.2	Data Collection and Processing					▓	▓						
3	**ASSESSMENT**												
3.1	Analysis							▓	▓				
3.2	Report								▓	▓			
4	**DISSEMINATION**												
4.1	Paper & Communication Output										▓	▓	
4.2	Feedback Seminar											▓	▓

Assessment: Quantitative and qualitative results are analyzed and, after feedback and validation from governments and country teams, results are reported in a standardized Impact Assessment Report, which includes the indicators of interest and provides answers to the questions identified in the

Figure 3. Example Timeline for Selected Projects

TENTATIVE TIMELINE	IMPLEMENTER	2018			2019												2020								
		OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
Mali (RFP)	IFPRI																								
Lesotho (SADP)	RIA																								
Bolivia (ACCESOS)	RIA																								
Mauritania (PASK II)	C4ED																								
PNG (PPAP)	C4ED																								
Kenya (UTaNRMP)	IFPRI																								
Nicaragua (NICADAPTA)	C4ED																								
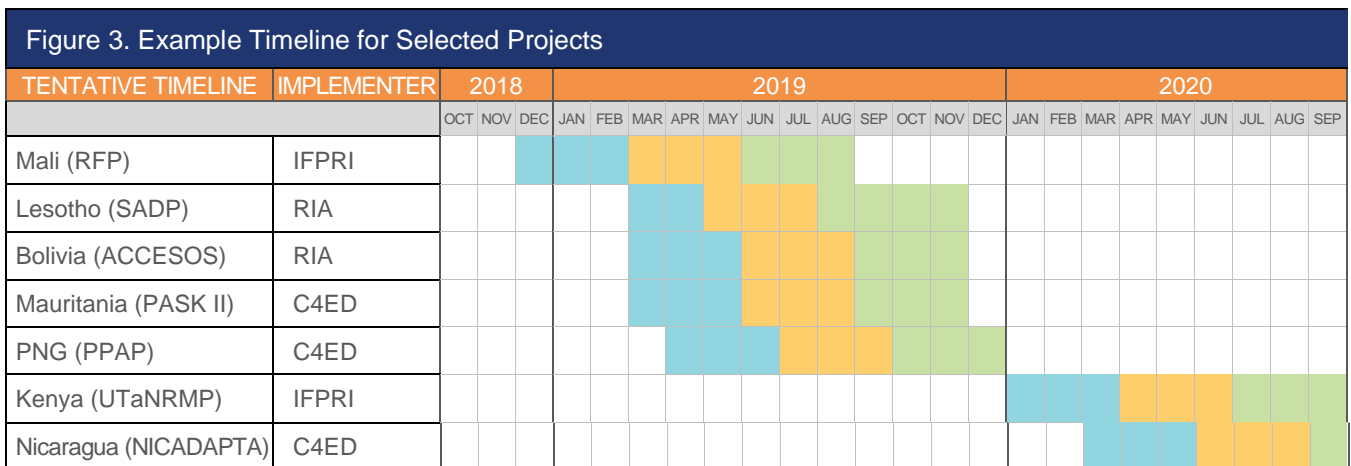
Impact Assessment Plan.

Dissemination: Ancillary products (briefs and infographics) are developed to accompany the more technical Impact Assessment Report and dissemination activities are undertaken.

The whole cycle is expected to last about 12 months on average, though this varies significantly depending on the availability and quality of data for sampling design, which is part of the planning step. Figure 2 shows a sample Gantt chart of specific IA activities and deliverables over a 12-month period. It should always be taken into account that the time assigned to each activity might be revised in light of specific constraints and delays outside the control of the IA team. For instance, in the case of the IA of the IFAD10's DECOFOS project in Mexico, both the household and the community surveys were supposed to be administered by September 2017, but due to unforeseen circumstances such as the earthquakes that struck Mexico's southern states in September 2017, data collection was delayed and instead completed between December 2017 and January 2018.

Impact assessments are conducted over three-year periods that overlap with IFAD's replenishment periods. The Eleventh Replenishment of IFAD's Resources (IFAD11) covers the period of 2019-2021, and RIA has selected 24 projects that are closing during IFAD11 for Impact Assessments using a protocol discussed in detail in the next section. The results of the whole IA cycle are expected to be produced before the end of 2021 such that RIA IA outputs can be aggregated to produce corporate Impact Analysis report and subsequently presented to IFAD's executive board in 2022 Figure 3 shows an example timeline of a set of selected projects for IFAD11 IAs.

The rest of this IA manual is organized into **four** core sections representing the phases of an impact assessment: planning (Section 2), implementation (Section 3), assessment (Section 4) and dissemination (Section 5). These are complemented by an annex, which is aimed to guide analysts for hands-on implementation with details on RIA document repository and file/folder organization.

2. PLANNING

The first step of the IFAD IA cycle consists of a planning phase characterized by a crucial set of activities that have to be carried out to select projects from the IFAD's portfolio of projects closing in that replenishment period to be evaluated through an impact assessment by RIA.

2.1   Screening and Contact with PMU and CPM

Screening Process

The starting point of the planning phase is represented by a process of **screening** and **selection** of projects for which the impact assessment will be conducted. The results of individual impact assessments are then aggregated in a meta-analysis and projected to the whole portfolio for corporate reporting. These initial activities are done in collaboration with IFAD's regional divisions and OPR that are requested to identify and

select potential countries and projects. A first screening is done through a participatory approach with five regional divisions based on the following characteristics:

- Disbursement rate
- Timing of the project (ideally at least 1.5 years before project completion)
- Geographical and thematic representation of IFAD's project portfolio

After a first screening, the selection process is conducted on the basis of the following key criteria:

- Technical feasibility to do a rigorous impact assessment
- Learning potential
- Quality of monitoring and evaluation (M&E) data
- Number of beneficiaries
- Type of interventions
- Buy-in from local governments and IFAD project teams;
- Research being conducted on the project to avoid overlap

Box 1. From Projects to Aggregate Corporate Impact

The steps from the review and selection of projects to the aggregate corporate impact through RIA impact assessment activities

REVIEW OF PORTFOLIO → **SELECTION** OF PROJECTS → **ESTIMATES** OF IMPACT across selected projects → **META-ANALYSIS** and projection to the portfolio

Number and **types** of projects

Projects closing during replenishment period

At least 15% of portfolio

Using a selection protocol

X MILLION PEOPLE IMPACTED

In the case of IFAD11, RIA had started the process of identifying candidate IFAD projects for *ex-post* impact assessments in July 2018 as described in the IFAD11 IA project selection process, which can be requested from RIA (riamailbox@ifad.org). Box 1 is a stylized description of the approach employed to measure impact, where a selection of projects (around 15 per cent of the entire IFAD portfolio of projects closing during a replenishment period) representative of the IFAD portfolio is used as a means to measure corporate impact. Towards this end, the methodology applied allows to attribute IFAD's impact at the corporate level, namely providing an estimate

of aggregate impact for the corporate indicators that are laid out in the IFAD
[Development Effectiveness Framework.](#)

IFAD's Strategic Framework 2016-2025 has one corporate goal and three strategic
objectives (SOs) as follows:

Goal: Increase economic mobility

   I.   SO1: increasing the productive capacity of poor rural people

  II.   SO2: increasing their benefits from market participation

 III.   SO3: strengthening the environmental sustainability and climate resilience of
their economic activities

These strategic objectives are in turn mapped to the Sustainable Development Goals
(SDGs), particularly SDG1 and SDG2 as described in Table 1

Table 1. Impact Indicators and Associated Targets		
Impact indicators	EG/SO	SDG
Number of people experiencing economic mobility	EG	SDG 2.3 SDG 1.2
Number of people with improved production	SO1	SDG 2.3
Number of people with improved market access	SO2	SDG 2.3
Number of people with greater resilience	SO1	SDG 1.5
Source: Report on the IFAD11 Results Management Framework, September 2017.		

The ideal approach to estimating attributable impact would be to design *ex-ante* IAs
that start prior to project implementation and collect baseline data as well as endline
data from the same sample. *Ex-ante* design, however, has proven to be very
challenging in IFAD projects given the long project design cycle, long project durations
and the need to keep the control group non-contaminated for the entire project
duration. These are exacerbated with the lack of capacity for implementation (e.g.
randomised phased in design) in country teams. A small number of ex-ante IAs are
implemented by RIA, which are increasingly driven by demand from project and
regional teams as IAs are appreciated within IFAD. Nevertheless, it is important to note
that almost all impact assessments for corporate reporting use an *ex-post* design.

Either way, in order to identify the impact of a particular project, or a "treatment", it
would be ideal to compare beneficiaries of the project and those same beneficiaries in
the absence of the project. Since beneficiaries cannot be simultaneously in and out of
the treatment group, it is necessary to identify a control group that is comparable to the
treatment group. The feasibility of identifying such a control group for the IA (i.e.
evaluability) is one of the criteria used in the screening and selection process.

At the end of this process, a sample is selected, and a validation exercise is conducted in collaboration with Regional Directors, Country Directors, Portfolio Advisors and Regional Economists of IFAD. Finally, the selection of projects for impact assessments is confirmed by each Regional Director to ensure full support of the regional division.

Box 2. Attribution Versus Contribution

When setting up an impact assessment, a clear understanding of the purpose of the exercise important. The purpose could either be to establish attribution, which means being able to establish a causal link between observed changes in outcomes and a specific project intervention; or simply, contribution, the extent to which a specific intervention on the ground has helped to achieve or was part of changes observed in the outcomes of interest.

Attribution: One can credibly claim that the impact of an intervention is due to the project's intervention only. In an ex-post IA framework, attribution can be established only through a carefully designed sampling frame which hypothesizes the existence of a counterfactual and shows the causal impact of the project by collecting detailed data from a sample of both beneficiaries and a comparison group.

Contribution: One can only establish if the outcomes or impacts of interest improved during the intervention period but cannot infer any causation between the project and the outcome or impact observed. To establish contribution, a sample of only beneficiaries of the project is sufficient.

Box 3. Q&A of Measuring Impact

QUESTION: How do we know if the impact we observe is caused by, i.e. attributable to, the project instead of other factors?

ANSWER: We need to find out what would have happened in the absence of the project, by creating the so called counterfactual.

QUESTION: How do we create the counterfactual?

ANSWER: Experimental methods randomly select the households that will benefit from the project (treatment) and those that will not. In this case, impact can be measured simply as the difference in outcomes between the two groups.

In other cases, beneficiaries are not randomly selected from an "eligible population", so an experiment is simulated by constructing a control group (i.e. households that do not benefit from the project) that is as similar as possible to the treatment group, using techniques such as propensity score matching. Impact can then be estimated as a difference in outcomes between the two groups using a wide range of quasi-experimental methodologies.

```
              ┌──────────────────────┐
              │  ELIGIBLE POPULATION │
              └──────────────────────┘
           ┌─────────────┴─────────────┐
  ┌──────────────────┐      ┌──────────────────┐
  │ TREATMENT GROUP  │      │  CONTROL GROUP   │
  └──────────────────┘      └──────────────────┘
```
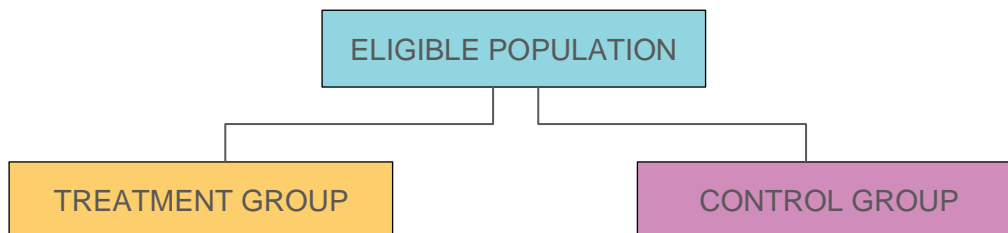
Once a project is selected for an impact assessment, RIA starts preparatory work by conducting a **desk review** of project documents and reports and interacting with the project team to gather a better understanding of the intervention, its logic and implementation framework. Key documents in this regard are the Project Design Report, Supervision Mission Reports, Mid-Term Report and Project Completion Report (if available). Project Documents by Country are available on the XDesk and accessible to IFAD users only.

Box 4. Theory and Practice of Impact Assessments in Developing Countries
GOAL: To enhance in-country capacity to conduct IAs of projects, especially of governments and development institutions involved in monitoring and evaluation (M&E). **OBJECTIVES:** To strengthen in-country capacity to conduct rigorous IAs of rural development projects To learn lessons from experience gained on implementation of rural development projects related to impacts and reachability of targeted beneficiaries as well as on potential barriers encountered by projects To measure the effectiveness of rural development projects through impact assessments in different countries and regional contexts, which contributes to the global knowledge on what works in rural development and what does not

First Contact with CD and PMU

While discussing the impact assessment with the project team, the following key actors are expected to be included in the correspondence and frequently kept informed about the status of the activities:

- Regional Director
- Portfolio Advisor
- Regional Economist
- RIA IA focal person from each regional division
- Country Director
- Programme Officers and Country Programme Officers

As a first step, an introductory email is sent to the Country Director and Programme Management Unit providing a brief overview of the purpose of the impact assessment as well as to start planning for the scoping mission. In case the IA for a given project is financed through a grant, the RIA grant manager sends this first email and is kept informed of the status of the IA as well as all issues that might arise.

2.2 Scoping Mission

After concluding the screening and preparatory work, a **scoping mission** is undertaken in the country and a series of meetings are organized with all relevant project stakeholders with the following objectives:

- Finalizing the ToC of the intervention
- Identifying the main research questions
- Defining the key indicators to be measured to assess the IFAD project's impacts
- Holding a capacity development workshop on the methodology
- Drafting the approach for the sampling frame

At the end of the mission, RIA and the project team agree on a work plan, which helps the RIA team to finalize the impact assessment plan. Sometimes, in order to ensure high-level buy-in and collaboration from the country's implementing ministry, official messages in the form of Protocol Letters are sent to formally inform government representatives about upcoming impact assessment activities. If this is the case, the RIA officer prepares the letter (signed by the RIA director and the country director) to be sent to selected stakeholders in the country before the scoping mission.

Preparing for the Mission

The **Terms of Reference (ToR)** for the scoping mission describe the purpose and organization of the mission. The document is prepared by the RIA officer who is responsible for the IA cleared by the RIA director and shared with the CPM and other relevant individuals in the country before the onset of the mission. TORs are critical for defining the following elements:

- Objectives, activities and expected results
- Information related to the mission, such as travel information, accommodation, and contact details
- Agenda of the mission

The main **objectives of the mission** are often related to:

- **Assessing the feasibility** of the impact assessment, namely confirming that the characteristics of the project and its implementation are suitable for an impact assessment, as well as ensuring interest, engagement and buy-in from the PMU
- Discussing and validating the **project logic and ToC** of the intervention with the project team including the definition of key indicators to measure
- Holding a workshop on the methodology both as a capacity development tool for the PMU, as well as ensuring a good understanding of implementation needs (esp. the identification of a control group)
- Developing an appropriate **sampling and identification strategy** in collaboration with the project team
- Planning impact assessment activities of the project to be evaluated

Specific activities planned for the mission are as follows:

- Discussing the **purpose, relevance, and rationale for the IA** of the project with the IFAD country team, the PMU and other relevant stakeholders
- **Collecting information** on project implementation, deployment, timing, schedule of activities, geography, targeting, eligibility, and beneficiary selection criteria. This includes obtaining information and access to data collected by the project team and/or other secondary data available, and get recommendations on potential local data collection firms

- Discussing and working with the project team to elaborate the **sampling and identification strategy,** agree on the **timeline and work plan** of IA activities, as well as on the definition of **roles and responsibilities**
- Organizing **field visits** to engage with project beneficiaries to better understand participation decision, benefits obtained, mechanisms in place for project implementation and barriers encountered
- Meeting with project **M&E officers and representatives of the firm,** who carried out baseline and mid-line impact surveys, to assist on coordinating data collection and the exchange of any other information needed. RIA team members are expected to be able to access any data available before the end of the scoping mission, recognizing that receiving data would be harder and more cumbersome after having left the mission location.

Box 5. Do's and Don'ts during a Scoping Mission

DOs:
- During the scoping mission workshop, the participation of all relevant stakeholders is to be ensured
- The ToC, key questions, sample frame and IA plan is developed in a participatory manner

DON'Ts:
- It must be recognized that not all project components can be evaluated, especially if statistical power cannot be ensured
- The focus is expected to remain on the main areas of interventions to prevent wastage time and resources

---

During the scoping mission, through the activities previously planned, the following outputs are expected to be achieved:

- **Buy-in, engagement and support** of the PMU in conducting the impact assessment
- Discussion and validation of the project logic and ToC with the project team, and definition of the main impact assessment questions and related indicators to be measured
- Discussion and validation of possible strategies to identify project's beneficiaries and selection of a control group in collaboration with the project team
- Access to available data from the PMU and other relevant stakeholders in order to facilitate the selection of a control group and to build the sampling frame; linkages with the national statistical institute (or other relevant institutes) to ensure access to secondary data
- Good understanding of the intervention delivery mechanisms, including participation decision, benefits obtained, mechanisms in place, potential unintended impacts and spillover effects, barriers and difficulties

encountered that are instrumental to a proper and effective design of survey tools for the unit of analysis

- Discussion and agreement on the impact assessment plan with the IFAD country team, the PMU, and other relevant stakeholders. This includes the timeline and work plan of IA activities to ensure that all products are shared and delivered on time, as well as the definition of roles and responsibilities of the various stakeholders involved in the impact assessment

Presentation of IA Methodology

In addition to the aforementioned activities, the preparatory work for the scoping mission consists of adapting presentations and drafting the ToC to be discussed during the meetings organized with the PMU and relevant stakeholders, aimed at ensuring their interest and buy-in. These meetings also allow RIA team members to understand project components, activities implemented and main characteristics of different interventions in detail.

RIA Team members typically **presentations the IA methodology and the main steps of the impact assessment cycle at a workshop** attended by all relevant stakeholders. A participatory discussion is facilitated in these workshops to define the ToC of the project, including main research questions and key indicators to be measured through the impact assessment. The interaction between various stakeholder groups is generally very valuable and facilitates the understanding of the beneficiary selection process, the details of implementation and the main challenges faced in executing the project. In addition, it also facilitates obtaining access to M&E data as well as contacts with government officials involved in the design the project.

Determining IA Feasibility and Stakeholder Buy-in

One of the main objectives of the scoping mission is to assess the feasibility of the impact assessment of the proposed project. Following the meetings with the Project Management Unit and the IFAD country office, the RIA team examines all the characteristics of the project to be evaluated to confirm that it is suitable for a rigorous impact assessment. At the end of the scoping mission, the interest, engagement and buy-in from the project team should be also confirmed to ensure collaboration during the implementation of all IA activities.

Timeline of IA Activities, Roles and Responsibilities

Upon return to office, RIA members who participated to the scoping mission write a Back-to-Office Report (BTOR), which highlights the key decisions made during the mission. The BTOR has to be prepared and sent to the RIA Director for **clearance, and then to the whole division and country team within one week after each scoping mission.** The final document includes the following information:

- Basic summary of the mission including IFAD mission team, project team, and mission dates
- Purposes and activities of the scoping mission

- Main contacts and persons met
- Short description of the project
- Summary of meetings and actions
- Main outcomes and steps forward

The RIA team also provides inputs to the **Aide Memoire** including the progress made, as well as agreed future actions and recommendations. This document is prepared by the IFAD country team and shared with the project team and main individuals working on the IA.

2.3 IA Plan

The **impact assessment plan** is the result of all previous activities and lays out the conceptual and technical contents and the timeline of the activities to be performed. It includes the following key elements that are needed to ensure a rigorous *ex-post* IA of a given project:

- A thorough description of the logic and flow of the project's **ToC**
- The main **IA questions** emerging from this logic
- The **indicators** to be measured
- The overall identification strategy and the methods to be used to construct the **counterfactual**
- A description of the qualitative and quantitative samples
- The **survey instruments** to be used as well as any complementary data collection to be undertaken
- The budget, key deliverables, and **work plan.**

The ToC (ToC) specifies the inputs, activities, outputs and outcomes of the project, and identifies the channels through which the expected impacts are to be generated. It helps the team identify impact indicators to be measured (some of which are identified in project log-frames) on direct beneficiaries. The IA plan also needs to include whether significant secondary or indirect (spill over) effects are expected, and if so, how they can be incorporated in the sampling frame. It also needs to identify any relevant heterogeneous impacts on different sub-groups of beneficiaries or under different circumstances that need to be taken into account during questionnaire and sample design.

IA Plan Templates by Region are used depending on the country, where the IA is being conducted. The IA Plan Design Guide may be consulted when developing the plan.

IA Plan Outline

The IA plan is prepared with the purpose of providing guidance to all the stakeholders who need to know details of an impact assessment. The IA plan includes both technical and practical elements so that the proposed approach to implementation of an IA to facilitate ease of understanding.

Each IA plan outline includes an introduction, a ToC, impact assessment questions, an IA design, as well as a clear budget and timeline. In particular, the timeline is a living document that is continuously updated as plans are finalized.

The **Introduction** contains:

- project name and timing of implementation
- objective of the document
- justification for undertaking IA for the project, including relevance for IFAD, government, and development in general
- main project components and their rationale.

The section on **ToC and main IA questions** offers a description of the logic and flow of the project's ToC and questions emerging from its logical framework. This section includes:

- a **clear description of the development problem** that is sought to be resolved, noting the literature on this issue
- a **well thought, clear and structured ToC** that clearly indicates logic of the intervention, expected activities, outputs, outcomes, impacts (including a narrative and a diagram representing the theory), including citations from the literature of the anticipated success of such an approach
- a **description of the target population** with a clear identification of criteria for selection of beneficiaries
- the identification and discussion of **possible unintended impacts**
- the existence of any possible **spillover effects** that may result from the project
- **evaluation** questions the IA is trying to answer, including citations from the literature on previous attempts to answer these questions
- a clear link to IFAD Strategic Objectives (SOs) which are aimed at increasing poor rural people's productive capacities (SO1), market participation (SO2), and resilience (SO3). Links with Sustainable Development Goals (SDGs) can be also highlighted when relevant

The section on **Impact assessment design** provides a description of the overall identification strategy and methods used and specifically includes the following information:

- a description of overall approach and methodological strategy to assess impacts
- the specific manner in which the treatment and control groups are expected to be created and its link to project design
- identification of potential spillover effects and, if relevant, a description of their measurement and ways in which contamination of the control group could be avoided
- a description of the quantitative, qualitative and any other complementary methods and sources of data and information that will be used to assess

impact, clarifying the manner in which the different methods will be combined to create a reasonable approach to assess impacts
- link to project monitoring and evaluation system
- the timing of IA activities to be performed and their relationship with project timing

The section on **Sampling and Data Collection** includes a detailed description of key indicators mapped to IFAD's goal, strategic objectives, and mainstreaming themes, as well as qualitative and quantitative samples and instruments, as follows:

- The sub-section on **key indicators** identifies and defines short, medium, long-term outcomes and impacts, as well as their indicators and data sources

- The **Qualitative Sample** part gives a description of:
  - the unit of analysis (individuals, groups, etc.) and the type of unit (key informants, focus groups, etc.)
  - a clear indication of the sample size needed for each unit of analysis, strategy, and approach for selecting the sample

- The sub-section on **Quantitative Sample** offers a detailed description of:
  - the unit of analysis (households, communities, provinces etc.) and any **clustering** of units
  - the sampling **strategy** and sample **frame**
  - the **power calculations** and the method for determining the **sample size**
  - a clear indication and distinction between **treated and control** and of the data sources used
  - the potential **attrition** and eventual **sample adjustment** needed

- The sub-section on **Qualitative instruments and method** includes the following elements:
  - questions or topics for focus groups
  - key issues or questions for key informants
  - information included in value chain analysis
  - approach to institutional mapping
  - other approaches
  - strategy for analysis (e.g. content mapping)

- The sub-section on Quantitative instruments and method includes the following:
  - questionnaire modules with a clear link to the ToC and to indicators to be measured

  - specific anticipated statistical method(s) for data analysis—equation(s) to be included if not straightforward, level of clustering or correcting of standard errors noted, impact to be estimated (ITT, ATE, ATT, LATE), methods (if distinct) for additional analyses such as spillovers and subgroup analyses noted, etc.

- **Complementary data** (e.g. GIS, climatic data) is expected to be included if directly contributing to impacts assessment analysis. In all cases RIA uses GIS for sample design, in which case they need to be included in the quantitative sample section.

The section on **Budget, Deliverables and Work Plan** provides practical information of the IA plan, such as:

- a detailed budget
- the key deliverables including responsibilities and link to standard project documents
- validation of results and dissemination plan
- identification and composition of IA team and responsibilities
- a detailed timeline in the form of a chart including timing of IA activities and budget for each period in question.

Elaborating the Theory of Change

One of the key aspects of conducting an impact assessment is the development of the ToC that begins during the scoping mission when the RIA team discusses the logical framework of the intervention with the project team. In fact, the ToC embedded in the project log-frame represents the starting point for a rigorous impact assessment (Figure 4). The TOC is defined as the causal chain that links project activities (inputs) to the results (outcomes and outputs) and impacts. The latter is spelled out within project design documents.

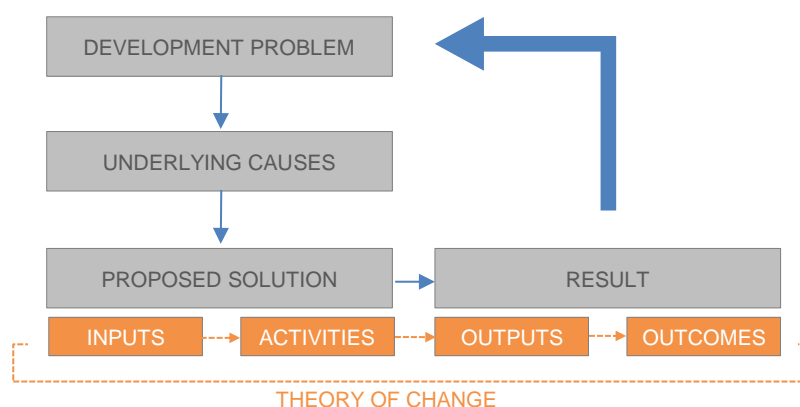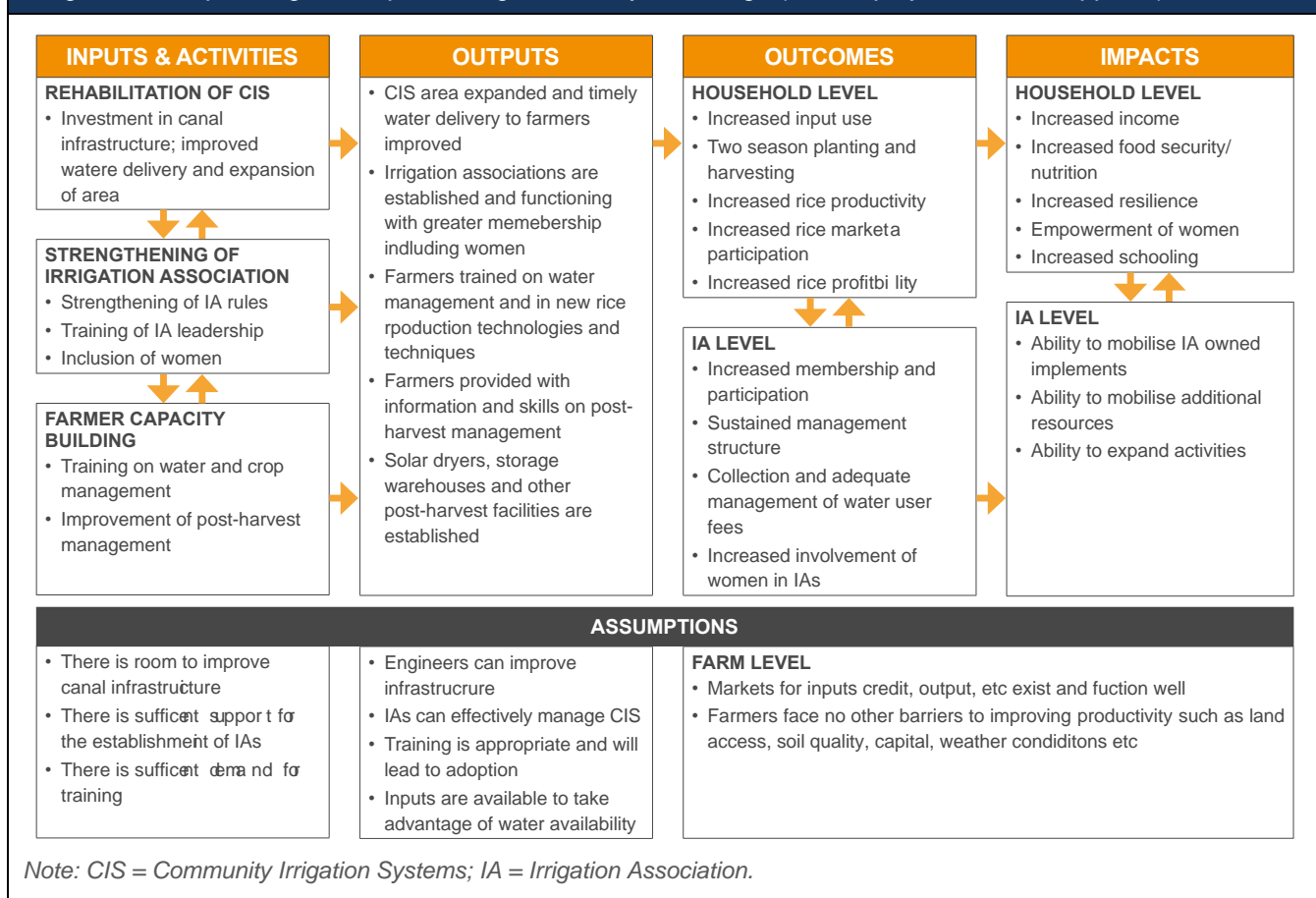Figure 4. Theory of Change as Part of Effective Projects

Figure 5. Sample Diagram Representing the Theory of Change (IRPEP project in the Philippines)

INPUTS & ACTIVITIES	OUTPUTS	OUTCOMES	IMPACTS
REHABILITATION OF CIS • Investment in canal infrastructure; improved watere delivery and expansion of area **STRENGTHENING OF IRRIGATION ASSOCIATION** • Strengthening of IA rules • Training of IA leadership • Inclusion of women **FARMER CAPACITY BUILDING** • Training on water and crop management • Improvement of post-harvest management	• CIS area expanded and timely water delivery to farmers improved • Irrigation associations are established and functioning with greater memebership indluding women • Farmers trained on water management and in new rice rpoduction technologies and techniques • Farmers provided with information and skills on post-harvest management • Solar dryers, storage warehouses and other post-harvest facilities are established	**HOUSEHOLD LEVEL** • Increased input use • Two season planting and harvesting • Increased rice productivity • Increased rice marketa participation • Increased rice profitbi lity **IA LEVEL** • Increased membership and participation • Sustained management structure • Collection and adequate management of water user fees • Increased involvement of women in IAs	**HOUSEHOLD LEVEL** • Increased income • Increased food security/ nutrition • Increased resilience • Empowerment of women • Increased schooling **IA LEVEL** • Ability to mobilise IA owned implements • Ability to mobilise additional resources • Ability to expand activities

ASSUMPTIONS

• There is room to improve canal infrastructure • There is sufficent support for the establishment of IAs • There is sufficent demand for training	• Engineers can improve infrastrucrure • IAs can effectively manage CIS • Training is appropriate and will lead to adoption • Inputs are available to take advantage of water availability	**FARM LEVEL** • Markets for inputs credit, output, etc exist and fuction well • Farmers face no other barriers to improving productivity such as land access, soil quality, capital, weather condiditons etc

Note: CIS = Community Irrigation Systems; IA = Irrigation Association.

As described in Figure 4, in order to be effective, specific interventions need to be designed to address a given development problem. Through a careful analysis of the underlying causes, a set of sources of the problem as well as potential solutions could be identified. The proposed solution is expected to be effective in targeting the underlying causes that represent a constraint for development, and at the same time, the available evidence is expected to be used to address the problem. The proposed solution of a development problem is made of a series of inputs and activities leading to specific outputs, with the expectation that beneficiaries respond in a way that leads to anticipated outcomes and foreseen impact. In the pathway from inputs, activities and outputs to outcomes and impacts, a project should have an underlying logic, namely a ToC describing the pattern from proposed solution to expected results.

Therefore, the ToC seeks to explain the logic of the intervention, its underlying causes, and the proposed solutions. In

KEY DEFINITIONS
WHAT IS THEORY OF CHANGE (ToC)?

It describes how activities are understood to produce a series of results that contribute to achieving the final intended impacts. A ToC can be elaborated either for a project, a programme, a policy, a strategy or an organization.

particular, it shows how the inputs provided and activities implemented through the project are associated to particular outputs, as well as how expected outcomes implied by project's outputs will lead to expected final impacts. A well thought, clear and structured ToC clearly indicates **logic of intervention**, **expected activities**, **outputs**, **outcomes**, and **impacts**, including a narrative and a diagram representing the theory as presented in Figure 5. Understanding impact implies the need to know the context where the intervention took place. Essentially, this means understanding stakeholder's needs and views regarding what the project is expected to deliver on the ground.

Conducting a Literature Review

The ToC and main IA questions section also contains citations from the literature on the development issue that is sought to be resolved through the impact assessment. The Agriculture Research Digest collecting research papers published on different journals can facilitate the review of available literature relevant for the projects to be assessed. For all articles, the main information such as title, suggested citation, highlights abstract, and conclusion are reported. A RIA article code is attributed to each item in the list and the paper can be easily downloaded using the relative link. Whenever new and useful references are used in the review, they should be added to Digest file following the template included in the relative excel spreadsheet.

Sample Design

The sample design is the strategy for selecting the units to be surveyed such that the sample is representative of the desired population. This step is very critical for ex-

post impact assessments, and sometimes takes longer than the survey implementation. Since it is not usually possible to survey the whole desired population due to budgetary reasons, accurate sampling of a subset of the outreach population can reduce costs and provide representative information from which one can generalize for the total population. Accurate sampling requires a **sample frame** or lists of all units in the population of interest. The unit is usually the individual or household from which one is interested in collecting data. A sample frame for a household survey would include all the households in the population identified by location. Note that units can also be groups, producer organizations, cooperatives, and other entities. Lists of such units need to be maintained and updated throughout the duration of the project to guarantee a reliable sample frame. The sample design is one of the crucial activities of a survey to be included in the IA plan. The first step of this activity is to identify the population of interest from which the sample to be surveyed will be drawn.

Identifying the eligible population prior to designing a survey can be a challenge: while the latter can be easily determined in an *ex-ante* framework, this becomes challenging in an *ex-post* framework. **The main challenge is to find a counterfactual group namely, a group of households, farmers or organizations that are similar to the treatment group that is receiving the project intervention. They represent what the treatment group would have looked like if they had not received the treatment.**

Box 6. Random Sampling

Random sampling is the only way to get a representative sample of the population of interest. We typically cannot survey every project beneficiary and so must define a sample to survey. A random sample is defined as a sample where each individual member of the population has a known, non-zero chance of being selected as part of the sample. Each individual is chosen randomly and entirely by chance. We expect that random selection will produce a sample that looks identical in characteristics to the larger group of interest (say, all beneficiaries).

As mentioned, in cases where the intervention was implemented at a level other than the household, the sample could be drawn at different levels. For example, if farmers' cooperatives are the first entry points for the project, a first sample of cooperatives could be drawn within which a second sample of members/households are drawn. Such a set-up would allow for a comprehensive analysis of project impacts.

When conducting impact assessments, one of the most important activities is planning a sampling strategy to guarantee enough statistical power in the analysis of project impacts, as well as in the identification of heterogeneous effects. Access to census data as well as use of first and second-hand information provided by governmental agencies and project units are extremely useful for sample calculations. Statistical power calculation is performed to establish the number of households to be surveyed for assessing project impacts using the Power formula (World Bank, 2007), which is usually adopted when conducting IA. Although sample size might be limited by budget or implementation constraints, a meaningful target power ranges between 80% and 90%.

RELEVANT READINGS

BetterEvaluation, 'Understand Causes', web page, BetterEvaluation

Gertler, P. J., et al. (2016). Impact Evaluation in Practice, World Bank, Washington, D.C.

Khandker, S. R., et al. (2010) Handbook on Impact Evaluation: Quantitative Methods and Practices, World Bank, Washington, D.C.

Shadish, W. R., et al. (2002) Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Houghton Mifflin Company, Boston, pp. 103–243

> **Box 7. Heterogeneous Impacts - Stratification**
>
> - In the case of agricultural projects, different interventions might affect project beneficiaries in a different way. When this is the case, it might be worth measuring impacts on different types of beneficiaries to identify heterogeneous outcomes separately.
> - Stratification is recommended to obtain statistically valid impact estimates by subgroups of interest, such as region, crop producers, livestock farming, and those with access to irrigation.
>
> The sample is stratified to guarantee representativeness within different groups in a given population (strata)

A standardized DO File to compute the sample size is available using the Stata command '*power*' which can also account for intra-cluster correlation.[1] Furthermore, this calculation allows to obtain the sample size required for statistical analysis, although it might be increased to account for some observations that will be dropped due to problems during field work or when processing data. In any case, this observation loss should be minimized to the extent possible to avoid the introduction of a source of potential bias into the sample and subsequently in data analysis. The experience of project and country teams might provide a sense of the average number of observations that will be dropped during data collection, as well as in understanding potential ways to reduce this number.

KEY DEFINITION

Random assignment should not be confused with random sampling. Whereas RANDOM SAMPLING refers to how a sample is drawn from one or more populations, RANDOM ASSIGNMENT refers to how individuals or groups are assigned to either a treatment group or a control group.

Once the number of observations is obtained from the formula, this number is doubled to sample comparison and treatment groups. Both groups should be of equal size. Ultimately, sample size is an **educated guess**, and it works only if the following conditions hold: *i*) the study samples come from the same or similar populations to the pilot study populations; *ii*) the population of interest is not changing over time; *iii*) the difference or association being studied exists.

---

[1] *All files referred to in the rest of this manual can be requested from RIA by writing to:* *riamailbox@ifad.org*

Box 8. Power Calculation Formula

The Power calculation formula allows to determine the total sample size (N)

$$N = \frac{4\sigma^2 (Z_\alpha + Z_\beta)^2}{D^2}$$

D: Impact on the outcome variable measured as the difference in means

σ: Standard deviation of the outcome variable

Z_α: Critical value of a confidence interval (two tail test=1.96)

Z_β: Critical value of the statistical power (two tail test=1.28)

The following corrected formula takes into account intra-cluster correlation due to the fact that units within the same cluster tend to be similar:
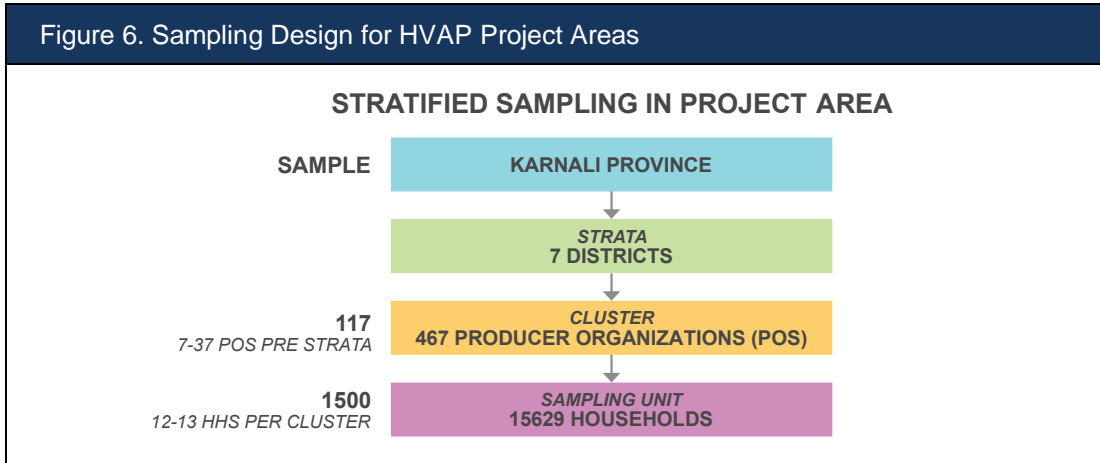
$$N_{Corrected} = N[1 + \rho(m - 1)]$$

ρ: intra-cluster correlation coefficient / m: number of units interviewed in each cluster

When ρ is high, more information are needed because additional observations within clusters provide less information since intra-cluster units are very similar.

The Stata command to compute N is sampsi <treatment value> <control value>, sd(<option>), whereas the one to compute N$_{corrected}$ and correct for clustering is sampclus.

Once the optimal size has been obtained through the power formula, the sampling strategy adopted usually depends on project's characteristics and units of analysis. For example, households can be part of producer organizations, cooperatives, water user associations, credit/saving groups, or other entities that should be taken into account when selecting the sample. When doing IAs, RIA usually employs a multi-stage stratified sampling to assure a representative sample at the geographical as well as at the group level Figure 6. Sampling Design for HVAP Project Areas presents the sampling strategy for the **High Value Agriculture Project in Hill and Mountain Areas (HVAP) for Nepal** covering 15,629 households (101,959 individuals) belonging to 467 Producer Organizations (POs).

Figure 6. Sampling Design for HVAP Project Areas

STRATIFIED SAMPLING IN PROJECT AREA

SAMPLE	KARNALI PROVINCE
	STRATA **7 DISTRICTS**
117 *7-37 POS PRE STRATA*	*CLUSTER* **467 PRODUCER ORGANIZATIONS (POS)**
1500 *12-13 HHS PER CLUSTER*	*SAMPLING UNIT* **15629 HOUSEHOLDS**

In the first stage, the project area is stratified into seven sub-populations (districts) and list all Producer Organizations (POs) covered by HVAP in each district (strata). The optimal project sample size (i.e., 1500 households) and the minimum number of sampling units per cluster is used (i.e., 12/13 units) to determine the required **number of treated clusters**, that is dividing 1500 by 13 gives us the cluster sample size of 117, after rounding. With 467 clusters in total, the cluster sample represents 25.05% of the cluster population. To assure proportional representation of all clusters in the final sample, 25.05% of clusters (POs) from each strata (district) are samples by using simple random sampling with proportional allocation. As shown in Table 2, this exercise gives the distribution of the total of 117 project clusters across project strata.

To ensure sample balance, the **number of control clusters** is expected to be exactly the same, i.e. 117 clusters from the project areas are selected for the study. In the second stage, first all the households in the selected clusters are listed. Then,

KEY DEFINITION

Statistical power refers to the probability of detecting the impact of a specific intervention. To conduct power calculations and calculate the required sample size for an IA, assumptions are made regarding the expected effect size, the statistical significance level and the intra-cluster correlation.

the number of households to be sampled from each strata (district) are calculated based on the number of sample households per cluster. Households from each selected cluster are then randomly selected. As the required sample size (1,500) is not an exact multiple of cluster sample size (117), at least 12 to 13 households per cluster are sampled to meet the required sample size. The sampling weights – an inverse of the probability of a sample unit to be selected in the final sample- are calculated as the final step.

Table 2. Sampling Frame: Districts, Villages and Households						
DISTRICT	NO. OF VDCS	NO. OF POS	NO. OF HHS	PROJECT SAMPLE		
				VDCs	POs	HHa
1. Achham	8	26	928	4	7	88
2. Dailekh	28	63	2,097	12	17	217
3. Jajarkot	12	60	2,826	8	15	192
4. Jumla	19	60	1,466	6	15	192
5. Kalikot	13	62	1,811	7	15	194
6. Salyan	10	45	1,176	7	11	139
7. Surkhet	36	151	5,354	22	37	478
Districts	126	**467**	**15,658**	**76**	**117**	**1,500**

The final sample is expected to adequately reflect the underlying heterogeneity in project beneficiaries as it includes project beneficiaries from each district. For the control group, the same number of households outside of HVAP areas, but within the same district are sampled. At the same time, counterfactual clusters are chosen based on propensity score matching such that for each district, best-matched clusters outside of project areas but within the same district are sampled. The number of project and control households and clusters are exactly the same with a total of 117 clusters and 1,500 households from 76 village development committees (VDCs) or villages from both project and non-project areas. Based on the number of clusters to be sampled from each district, two sets of project clusters are produced by randomly selecting the required number of clusters from each district twice. Project clusters in each random set are accompanied by three potential control clusters chosen by using the propensity score matching approach. **For each randomly selected project cluster, the best matched control clusters are then identified with help from the PMU and HVAP social mobilizers who work in the field**. In this selection process, social mobilizers organize a local meeting in each selected cluster to collect input from village leaders and other individuals, if needed. Following the adopted RIA methodology, the sample selection DO file for HVAP offers an example on how to sample for IAs.

In the IA Plan of Plan VIDA-PEEP to Strengthen the Capacity of Communities and Families Living in Extreme Poverty in Cochabamba and Potosí in Bolivia, a two-stage stratification approach was used assuming that beneficiary households were distributed in communities, starting from the optimal sample size obtained. The sample was expected to be representative of the geographical spread of the project and as such, the first level of stratification was at municipality level. In this first stage, the total number of treatment communities to be surveyed was distributed across project's municipalities proportionally to the number of beneficiary communities in each municipality with respect to the overall beneficiary communities. The second strata was represented by the community status in terms of being leader or non-

leader (satellite)2. This second level of stratification ensures representativeness of both groups of beneficiaries therefore allowing to account for potential heterogeneity of intervention effects. In order to do so, once the number of communities to be surveyed in each municipality has been determined in stage one, this number was split into leader and satellite communities proportionally to their relative weight within the same municipality. Treatment communities to be sampled were randomly selected from the list of leader and satellite beneficiary communities.

As described in the IA Plan of the Guangxi Integrated Agricultural Development Project (GIADP) in China, the activities under the community infrastructure component and the agricultural marketing support activities were delivered to all administrative villages (AVs) covered by the project. In addition, in this case, the sampling strategy to select the treatment and the comparison AVs followed a two-stage stratification approach: (1) by county, and (2) by level of project intensity. First, the Provincial Project Management Office (PPMO) provided with a complete list of AVs (both project and non-project). From this comprehensive list of AVs obtained from the M&E system, the AVs in each county were ranked by the distribution of project activities and divided this distribution into quartiles. Then, a number of project AVs were randomly selected and stratified by the level of project activity intensity implemented in each AV. After obtaining a randomly selected list of project AVs to sample, the propensity score matching (PSM, with five nearest neighbours and with kernel) was run to come up with a tentative list of non-GIADP AVs which could serve as the counterfactual group (comparison) for the GIADP AVs. The PSM was conducted separately for each of the eight counties to ensure that GIADP AVs can only be matched with non-GIADP AVs within the same county. The purpose of the PSM was to ensure that households in GIADP AVs and non-GIADP AVs were similar in terms of observable characteristics available in the data provided by the project's M&E system. For matching project and non-project AVs, variables such as number of natural villages, female and minority population, cultivated area under paddy rice, hectares of dry area were used. As the project targeted AVs mainly based on poverty levels, the variables include the share of households belonging to either B or C category along with other AV-level characteristics. To ensure the appropriateness of the selected non-GIADP AVs as the counterfactual, the PPMO and CPMO staff members were consulted to help validate and select the non-GIADP AVs to be included in the final sample.

---

2 *The Plan VIDA implementation strategy supported the formation of ad-hoc constituted groups of communities (Grupos Zonales) for the formulation of project's plans. Each Grupo Zonal was constituted by a leader community and other non-leader (satellite) communities. Understanding the elements that drove the designation of leader communities would have been very important in order to account for potential differences between the two groups that could confound the attribution of project's impacts. Unfortunately this type of information was not available, but it was plausible to assume that leader communities might differ from satellite communities for a number of characteristics such as proximity to roads and better accessibility, closeness to markets, better infrastructure, but also higher level of social capital and organizational skills, etc. suggesting potential heterogeneity of impacts between the two groups. For this reason the sampling design of this IA ensured representativeness of both leader and satellite communities.*

Ideally, a complete list of all households in treatment and control areas should be used to randomly select households to be interviewed within each stratum. Such a list is however usually not available and **household listing** needs to be conducted. This is an imperative part of any household survey because having a current list of households in survey area allows to randomly select required number of households from the eligible population. For Instance, in the case of the High Value Agriculture Project in Hill and Mountain Areas (HVAP) in Nepal, household listing was done by the HVAP project a few weeks before the household survey. In project areas, the list of beneficiary households was updated and used to draw the treatment sample. In control areas, households considered eligible for HVAP in baseline (meaning they were theoretically eligible to receive HVAP interventions but were located outside the areas covered by the project) but that did not benefit from the project were identified and listed. Such list was used as the sampling frame to randomly draw the control sample. The listing should be done in collaboration with the PMU with the motivation that this kind of exercise is extremely helpful in improving the M&E system related to the project. RIA staff members should support and guide the project team members in conducting this activity providing them with a set of information to be included when doing the listing exercise. In some other cases, the listing is included in the TORs when the firm responsible for the data collection is hired.

Table 3. Key Indicators	
IMPACT DOMAIN	INDICATORS
EG: Income from producing, processing and marketing agricultural products (crops, livestock, fish, forestry), as well as other sources	Income by source Amount and source of sales Costs of production Consumption Nr. of hours of labour per type of activity
SO1: Production and productivity	Total production and yields for crops Total value of production Total value of production from livestock/household enterprise…etc.
SO2: Market access	Probability of selling in the market Access to inputs Share of total produce marketed Access to credit/insurance/information
SO3: Resilience	Income/livelihood diversification indices Nr. of on-and-off-farm activities Changes in consumption Nr. and amount of credit obtained Ability to recover from climatic and non-climatic shocks
MT: Food security & nutrition	Household Dietary Diversity Score (HDDS) Food Insecurity Experience Scale (FIES)

Data Collection and Key Indicators

RIA uses a mixed-methods approach to impact assessments combining quantitative (household and community or associations questionnaires) and qualitative data. The indicators of interest need to be defined carefully to measure the impact of project interventions on beneficiary livelihoods. The indicators need to be mapped to IFAD's economic goal (EG), strategic objectives (SOs) and mainstreaming themes (MT), as well as any project-specific domain. The IA plan gives information on the data collection method and offers a description of the impact domains and indicators that are generally constructed and used in the analysis as depicted in **Table 3**.

Table 4 shows output and outcome indicators selected for the IA of the DECOFOS project in Mexico as an example.

Table 4. Mexico DECOFOS' Key Indicators		
INDICATOR	DEFINITION	SOURCE
HOUSEHOLD LEVEL		
OUTPUT		
Investments in productive assets and infrastructure	# productive assets; value of expenditure in productive assets; value of expenditue in infrastructure	Household survey
Sustainable exploitation of natural resources	Ha under agro-forestry modules; # plant nurseries per Ha; # firewood saving stoves in use; # eco-toursim micro-enterprises started	Household survey
Access to financial resources	# and amount of loans obtained	Household survey
OUTCOME		
Increase in income	Net annual income per capita	Household survey
Diversification of economic activities	# income sources; income and labour diversification indices	Household survey
Dinamization of the local economy (increase employement opportunities)	# employees in household micro-enterprises and small-businesses; location of origin of the employees; type of employment contract; average wage per day	Household survey
Increase participation of women and youth in economic activities	# household micro-enterprises and small-businesses headed by women and youth; # women and youth empoyed in micro-enterprises and small-businesses	Household survey
Reduce youth migration	# young household members having migrated to urban areas or abroad	Household survey
IMPACT		
Reduce asset poverty and food insecurity	Value of food and non-food expenditure per capita; # and value of household assets; # meals per day	Household survey
Increase resilience	Exposure to negative shocks; adoption of agro-forestry and good environmental practices	Household survey
ENVIRONMENTAL LEVEL		
IMPACT		
Climate change mitigation	GHG emissions; carbon sequestration	Community GIS
Climate change adaptation	Ha under agro-forestry; # plant nurseries per Ha; # firewood saving stoves in use; Ha of natural resources conserved (e.g. Under eco-tourism businesses)	Community survey

Deliverables and Work Plan

The IA plan includes a detailed work plan that reports key activities to be carried out according to the timeline of the impact assessment as shown in Figure 7. Together with main activities, responsible RIA members as well as from country teams are expected to be identified and listed in this section. While the recruited firm and field supervisors are responsible for all the field work including recruitment of enumerators, pilot surveys, training activities and data collection, the RIA team

Once all the aforementioned information is gathered, the IA plan is prepared with the purpose of providing guidance to all the stakeholders about all relevant details of the impact assessment. **IA Plans for different countries** are available for consultation.

Figure 7. Sample Work plan												
IA CALENDAR NOVEMBER 2016–DECEMBER 2017												
ACTIVITIES	Nov 2016 —Jan 2017	Feb 2017	Mar 2017	Apr 2017	May 2017	Jun 2017	Jul 2017	Aug 2017	Sep 2017	Oct 2017	Nov 2017	Dec 2017
Finalize IA design & sampling strategy	▓											
Prepare survey instruments for qualitative analysis		▓										
Validate sample and conduct KII and FGD			▓	▓								
Recruitment of enumerators			▓	▓								
Develop quant survey instruments				▓	▓							
Enumerator training					▓	▓						
Pilot survey						▓						
Data collection							▓	▓				
Data cleaning								▓	▓			
Data analysis										▓	▓	▓
Draft report												▓
Validation of results												▓

2.4 Hiring a Data Collection Firm

Concurrently with the development of the IA plan, a **Mini-Tender** is launched to hire a data collection firm which is responsible to carry out all data collection activities under the close supervision of RIA. RIA has been given the opportunity to run its own mini-tenders and it is the only division in IFAD that is authorized to do so.

All vendors currently under Long Term Agreement (LTA) are listed in the UNGM portal and may be selected through the Mini-Tender. New vendors, from the next Generic Tender (GT) and any subsequent GT, are also listed in the system once they are under LTA. Access to the UNGM website is provided by the IFAD Procurement Manager.

The Generic Tender

A Generic Tender at IFAD identifies a number of competent contractors with experience in the development and administration of surveys to collect data needed for impact assessments in different countries. The contractors are selected on the basis of a generic tender and awarded a Long-Term Agreement (LTA), which does not have any financial commitment by IFAD on the minimum orders, but on the basis of "if and when required". A financial commitment by IFAD is only made to an LTA holder (contractor selected from this tender) once they participate and are selected as a result of a mini-tender with a specific and more detailed TOR. This process ultimately results in the issuance of a Work Order. Mini tenders are conducted when the specific requirement arises.

The contractors need to have expertise in field of data collection design and administration for impact assessments with preference to mixed method research designs, e.g. research designs that integrate both quantitative and qualitative (focus group discussions, key informants' interviews) methods of inquiry. They need to coordinate these activities with RIA. The scope of the services required include the performance of baselines and/or follow up surveys (at the appropriate level of analysis, namely, the individual, household, and community level). The end objective of these surveys is to evaluate the impact of a specific project in a specific location, measured through identified outcomes of success, which will be specific to the project's ToC, and will have to meet IFAD corporate requirements.

The **required services** include the following:

> ## KEY DEFINITION
>
> **The Purpose of IFAD Ex-Post Impact Assessments** is to ask policy relevant questions to generate an evidence base not only for country-specific dialogue and policy, but also for the international agricultural and rural development community.

- Preliminary analysis of context specific agricultural and poverty related indicators including descriptive statistics of the population and agriculture indicators, disaggregated at province or municipality where possible
- Design of adequate questionnaires, data entry programs, and training materials
- List of households in the primary sampling units (PSUs)
- Quantitative household level surveys
- Quantitative community-level surveys where relevant
- Qualitative surveys of community leaders and key respondents (for instance through focus group discussions, key informant interviews or other methods for qualitative inquiry)
- Geographic information system (GIS) services where appropriate.

Specifically, the services required include:

- Adaptation of the template IA questionnaire to the local conditions in consultation with IFAD
- Translation of the questionnaire into local languages including quality controls (e.g. reverse translation)
- Administration and analysis of a pre-test pilot of the adapted questionnaire in the field in the local language
- Adaptation of fieldwork manuals
- Adaptation and development of educational material for the training of the field personnel
- Data entry approach with adequate quality control systems
- Production of questionnaires, manuals, and final data entry program
- Organization plan of the fieldwork including all logistical arrangements
- Pre-selection of the field personnel
- Selection of the field personnel
- Training of the field personnel
- Production of the list of households/communities
- Performance of surveys according to the methodology agreed for the survey
- Periodic deliveries of the partial datasets to IFAD according to the agreed quality standards and adhering to the agreed format
- Implementation of the supervision program
- Production of consolidated datasets according to the agreed quality standards while adhering to the agreed format.

The contractors must meet the following requirements:

- Minimum experience with managing data collection surveys of development projects; (minimum 3 projects in the past 5 years) is necessary. More evidence of the projects in the Technical Proposal is an advantage
- Minimum five (5) years of experience in conducting surveys
- Ability to use electronic devices, such as tablets and/ or smart phones, for data entry and management
- Previous experience with both quantitative and qualitative data collection
- Previous experience with agricultural or rural development surveys is an advantage
- The contractors are expected to adhere to local laws while conducting the services outlined in the LTA (commitment letter in free format must be provided as part of the Technical Proposal)

The proposals are evaluated based on the content of the Technical Proposal and the following criteria:

- Details or profile of the vendor; company internal structure/organigram and CVs of key staff
- Description of the capabilities of the vendor to develop the requested activities, in particular demonstration of the vendor's experience in complex household and/or establishment surveys at the regional or national level. In addition, the vendor needs to indicate the country/ies of experience and expertise. Preference is given to vendors with proven expertise in mixed methods research designs. The relevant experience in this regard must be included in the proposal
- A description of the workflow related to the Services described in this Terms of Reference has to be provided as part of the proposal
- Reference letters or contracts from previous clients in the private, public and development sector
- Evidence or description of compliance with the requirements to the vendors mentioned above
- Any other documents that may support their proposal.

The contractor works in coordination with the principal investigator of the impact assessment (the designated IFAD-RIA Staff), consultants within the impact assessment team, the project management unit and IFAD country staff participating in each individual study.

The proposals must be limited to a maximum of 15 pages. The CVs and examples of previous relevant work must be attached as appendices.

Launching a Mini-Tender

Building a Mini-tender process requires the creation of an account in UN In-Tend Organiser Portal. The mini-tender procedure aims at hiring the **data collection company** that will conduct quantitative and potentially qualitative data collection for a specific impact assessment. The company can be **local**, **regional,** or **international**. The aim is to choose the best one in terms of **cost** and **quality**. Detailed information on how to build a mini-tender process in In-tend can be requested from RIA by writing to riamailbox@ifad.org.

The interested firms send to IFAD two different documents:

- Commercial Offer
- Technical Proposal

The necessary documents for launching the tender are:

- Mini-tender instructions providing useful information for the vendor, such as the opening and closing time of the tender, and due date for submission of bids.
- Mini-tender Terms of Reference to be prepared following the template available.

Often, vendors request clarifications or have questions on the details of the TOR. When requests and questions arise, they are expected to be forwarded to the Task Manager to respond. However, the name or contact details of the requester should be removed to ensure confidentiality. Responses are sent directly through the system. Depending on the nature of the query, the response is shared with all vendors.

Once all the bids are received, the proposals can be shared with the relevant staff, who then carry out the **technical and commercial evaluation**. If IFAD staff have queries on the contents of the proposals, they should request the RIA focal point to contact the vendors. RIA staff member should not contact a vendor directly, while a tender is in progress. This is a breach of protocol and could lead to the mini-tender being cancelled and sanctions against the staff member.

Panel members (PM) composed of technical experts carefully review the applications received, and shortlist few hiring data collection firms for an interview. The selection is made based on cumulative scores from **technical and commercial evaluations**. The **technical evaluation** will take into account the following criteria:

- Level of experience in carrying out qualitative and quantitative impact assessment surveys and in a context relevant to the requirements

- The firm presented a strong understanding of the requirements set out in the TOR and demonstrated capability in executing activities required
- The qualifications and experience of the proposed project staff
- The technical proposal presented complies with requirements set out in the TORs including tools used (i.e, type and capacity of tablets in terms of GPS requirements, translation process etc.)
- References provided by the contractors were valid and checked for positive feedback. It is good practice to check with at least a couple of references before contracting a firm

For each criterion, the firm may be considered excellent, good, or weak, getting a score depending on the assessment received.

The **financial evaluation** is based on a weighted evaluation of full costing of requirements provided.

In addition to the technical and financial criteria, the Evaluation Panel evaluates all technical proposals according to some **essential requirements** such as:

- The contractors' overall understanding of the TOR and requirements and that these were accurately reflected in the quality of the technical proposal.
- The budget properly reflected the requirements – the value for money principle was applied, with emphasis on quality.
- The contractor's proposal needs to score at least 35 in the technical evaluation.

If the response of any of the above is NO, the proposal is automatically disqualified from further evaluation.

Once the evaluation is completed, the following three main documents have to be sent to the Procurement Officer in the Administrative Services Division (ADM):

1. Memo
2. Work Order
3. Synthesis Report

The RIA front office assistants prepare the Purchase Requisition (PR) to be approved by the Budget Holder (BH). Once the assistant receives the Purchase Order (PO) the following working day, the Contract is ready to be sent to the Service Provider.

3. IMPLEMENTATION

3.1 Qualitative and Quantitative Tools

The implementation phase consists of developing the tools and instruments for data collection. RIA uses a **mixed-methods approach** in all of its *ex-post* impact assessments and develops both qualitative and quantitative tools at this stage. The qualitative research is used to triangulate results of the quantitative component and to give context to the interpretation.

Qualitative Tools

Qualitative research represents a very important approach for evaluating impacts of IFAD projects, although quantitative tools usually play a dominant role when conducting IAs. Qualitative investigations can either precede or follow the quantitative survey depending on the availability of funding and logistical feasibility. Qualitative analysis, if conducted before quantitative work, can provide information that can help the design of the quantitative tools as well as inform and complement findings from the quantitative analysis. In the presence of spillover effects, qualitative research could also shed some light on the impact while providing an understanding of the mechanisms through which spillover impacts are materialized.

In particular, the qualitative surveys aim at answering the question "what is driving change and how," which is particularly important in settings where there is the need to reconstruct the ToC and there is no baseline data. Qualitative tools include instruments for **Focus Group Discussions (FGDs)** and **Key Informant Interviews (KIIs)** consisting of semi-structured questionnaires. In some cases, a data collection firm is hired by the PMU to conduct the qualitative exercise with project funding before the quantitative survey. When it is not possible to carry out qualitative research ahead of the quantitative data collection, it might be useful to administer the qualitative survey at the same time as the quantitative survey (in a concurrent manner) in order to triangulate the results with the quantitative part of the impact assessment. Qualitative surveys allow gaining additional information related to project targeting, implementation, and the socio-economic and cultural project context.

FGDs are predetermined semi-structured interviews. The facilitator leading the conversation aims to generate the maximum amount of discussion and opinions among the participants within a given time period.

> ## KEY DEFINITION
>
> **Qualitative Research** can capture context specificities, by understanding characteristics of beneficiaries, processes through which changes occur, as well as why change did not occur. It allows to understand complex realities, mechanisms and processes that yielded impacts, such as pathways out of poverty, welfare enhancement and agricultural transformation (Rao and Woolcock, 2003).

Sample documents on Focus Group Discussions and Key Informant Interviews questionnaires for the Project for Rural Income Through Export (PRICE) in Rwanda can be requested from RIA by writing to riamailbox@ifad.org.

Quantitative Tools

All impact assessments require **quantitative** tools, both at the household and community or association level including detailed questions linked to the ToC, as well as all the variables needed for the analysis. In general, questionnaires should mirror the project ToC (outputs, outcomes, and impacts). The RIA household questionnaire comprises of **core modules** which include demographics, socio-economic

characteristics, and higher-level outcomes such as ownerships of assets (durables and productive) and food security (dietary diversity). It also comprises of modules that are project-specific and reflect the specific ToC of the project. It is in fact crucial that questionnaires include other variables beyond the core indicators, which are essential to contextualize the survey and used as control variables for the propensity score matching. Such variables are well suited to determine the validity of the counterfactual and can also describe other features such country-specific household typologies. Therefore, the household questionnaire template is conceived as an adaptable survey tool with a modular approach.

The RIA household questionnaire template has been developed to obtain a full understanding of the households' livelihood strategies and it is divided into 8 thematically specific modules. In each module, relevant outcome indicators can be measured using specific questions or can be derived from the aggregation of different questions. Certain modules and questions are not directly relevant for the outcome indicators but concern the measurement of the above-mentioned additional control variables, which are used for further analysis and disaggregation of outcome indicators by dimensions or sub-groups such as gender, age, indigenous people, sector etc.

The structure of the RIA household questionnaire template is depicted in Table 5.

Table 5. Main Structure of the RIA Household Questionnaire				
SECTION			SECTION	
A	HOUSEHOLD ROSTER		E	HOUSING
A-1	Demographics		E-1	Housing – Current
A-2	Literacy, Education And Disability		E-2	Housing – Past
A-3	Migration		F	ASSETS
A-4	Employment		F-1	Household Assets
B	PARCEL AND PLOT IDENTIFICATION		F-2	Agricultural Assets
B-2	Parcel Identification & Ownership		G	FOOD SECURITY/NUTRITION
B-3	Plot Identification And Use		G-1	Food Consumption - Hdds
C	AGRICULTURAL PRODUCTION		G-2	Food Insecurity Experience Scale - Fies
C-1	Annuals/Perennials Planting & Harvest		G-3	Food Shortages
C-2	Annuals/Perennials Post-Harvest Crop Uses		H	SHOCKS
C-3	Sales		I	SELF-EMPLOYMENT ACTIVITIES
C-4	Inputs		J	OTHER INCOME
D	LIVESTOCK		L	CREDIT
D-1	Livestock Characteristics		M	SAVINGS
D-2	Livestock Inputs		N	EXTENSION AND INFORMATION
D-3	Livestock Products		O	SOCIAL CAPITAL

MODULE A – HOUSEHOLD ROSTER collects information on household demographics and contains individual level questions on literacy, education, disability, migration, and employment. A household can be defined as a group of

individuals that eat together and share a common budget. This includes all members that live in the same dwelling, compound or close by. Members of a household do not need to be related by blood or marriage. If the household is polygamous, more than one spouse may be included if the spouse and associated household members live in same compound and share a common budget. Spouses of the household head that do not live together but share a common budget shouldn't be included, while those members who live elsewhere (for example, students at boarding school, head who have migrated temporarily) may still be included if they share the same budget. Note that sharing remittances does not constitute sharing the same budget. The definition of a household can vary depending on the local realities and it is important to verify it on the ground using the definition used in the Population Census or other local references. For example, in some contexts, a household is best defined as a group of individuals who share the "same kitchen."

MODULE B – PARCEL AND PLOT IDENTIFICATION is divided into sub-modules including parcel and plot use during the last 12 months and the last seasons. A parcel is usually represented by a contiguous piece of land and should not be split by a path of more than one meter in width. Parcels can be divided into plots. A plot must fulfil three conditions:

- a continuous piece of land (small paths between plots are possible)
- a unique crop OR a mixture of crops – differential intercropping scenarios are possible
- a uniform, consistent crop management system:
- a uniform and consistent use of inputs (incl. labour) on the plot
- a uniform and consistent water and land management system on the plot
- a single person making decisions on plot management (or, if more than one manager, they need to make decisions together).

A home garden is usually close to the homestead mainly used for cultivating crops for home consumption. It generally will not be larger than 0.1 Hectares (0.25 Acres/1,000 square meters/1,200 square yards/10,800 square feet). In case of extensive use of home garden, a specific section is added to the core modules of the RIA template questionnaire.

MODULE C – CROP PRODUCTION collects information on households' cropping activities, input use, land ownership and adoption of improved agricultural technologies.

MODULE D – LIVESTOCK records livestock ownership, stocks movement, and sales. Livestock names need to be contextualized to match livestock available in the local area. When fishing activities are relevant in the country where the IA is being conducted, household fishing activities have to be recorded with fish codes updated to reflect region specific fish species.

MODULE E – HOUSING records all the information relative to dwelling characteristics during the reference period as well as before the project start date.

MODULE F – ASSETS records ownership of durable and productive assets during the reference period as well as before the project start date.

MODULE G – FOOD SECURITY/NUTRITION allows the construction of different indicators such as dietary diversity scale, food insecurity experience scale and food shortages. The Household Dietary Diversity Score (HDDS) and Food Insecurity Experience Scale (FIES) questions are included in this module.[3] The examples of foods provided should be adapted to the local food context for each food group. The number of months with adequate food supply or food shortages is also included in this module where relevant.

Box 9. The PRO-WEAI

Building on the original WEAI (Alkire et al. 2013), the project-Level Women's Empowerment in Agriculture Index (Pro-WEAI) is a new survey-based index specifically designed for measuring the impact of agriculture development projects on women's empowerment, and it is used as a diagnostic tool for tailoring such programs to specific setting. The Pro-WEAI is calculated based on 12 equally weighted indicators mapped to three domains:

a. Intrinsic agency (power within)
b. Instrumental agency (power to)
c. Collective agency (power with)

	INTRINSIC AGENCY	INSTRUMENTAL AGENCY	COLLECTIVE AGENCY
1	Autonomy in income	Input into productive decisions	Comprised of group membership and membership in influential groups
2	Self-efficacy	Ownership of land and other assets	
3	Attitudes about IPV* against women	Control over use of income	
4	Respect among household members	Access to and decisions on financial services	
5		Workload	
6		Visiting important locations	

Although different from pro-WEAI, the RIA survey instrument allows the calculation of a roughly equivalent set of 12 indicators. In addition, the RIA survey instrument will contain expanded modules designed to measure collective efficacy within groups and agency over time.

* IPV: Violence Against Women Perpetrated by an Intimate Partner

---

For projects that aim to be gender transformative, the project-Level Women's Empowerment in Agriculture Index (Pro-WEAI) is also included in the questionnaire (see box 9).

In the case of Mali, one of the main objectives of the Rural Microfinance Program (PMR) was to facilitate access to finance by vulnerable women as well as to provide training on business skills to help them to navigate the loan process and to promote

---

3 FAO guidelines on HDDS: *https://www.fao.org/nutrition/assessment/tools/household-dietary-diversity/en/* and FIES: *https://www.fao.org/policy-support/tools-and-publications/resources-details/en/c/1236494/*

entrepreneurship. Hence, in addition to a specific section on microfinance, the quantitative instrument also includes some adaptations based on the project-level Women's Empowerment in Agriculture Index (Malapit et al. 2019) developed by IFPRI.

3.2 Data Collection and Processing

Once the data collection service provider is hired and the tools are finalized, the quantitative and qualitative **data collection** in the field can start after the **training of field personnel** and **pilot interviews** are carried out. Data collection is an essential step in the process of measuring project outputs, outcomes, and impacts. By measuring outcomes and impacts, it can be possible to recognize the effectiveness and value of IFAD projects and point any changes or improvements that may need to be made.

The quantitative data collection is conducted using the Survey Solutions software offered free of charge and developed by the Data group of The World Bank (co-financed by the World Bank, Bill and Melinda Gates Foundation and the Food and Agriculture Organization of the United Nations). Survey Solutions has proven to be a very powerful tool to ensure that high quality data is captured, either through face-to-face interviews or web-based interviews. In particular, Computer-assisted personal interviewing (CAPI) provides an alternative to traditional pen-and-paper interviewing (PAPI). Using CAPI, surveys can be conducted on low-cost Android tablets that adhere to the minimum standards required by Survey Solutions. The survey is pre-loaded onto a tablet and the interviewer reads the questions from the screen and enters the answers immediately into the device. Using Survey Solutions, data are transmitted to RIA in real-time through a dedicated software platform, while data quality checks are performed simultaneously by RIA and the data collection firm to allow for prompt feedback to field teams and quick resolution of problems.

RELEVANT READINGS

Gertler, P. J., et al. (2016) Impact Evaluation in Practice, World Bank, Washington, D.C.

Caeyers, B. et al. (2012). Improving Consumption Measurement and Other Survey Data through CAPI: Evidence from a Randomized Experiment." Journal of Development Economics 98 (1):19–33.

Fafchamps, M. et al. (2012). Using PDA Consistency Checks to Increase the Precision of Profits and Sales Measurement in Panels. Journal of Development Economics 98 (1): 51–57.

Enumerator Training

Before starting data collection, an enumerators' training is conducted. All the enumerators work in a team and report directly to the Data Processing Manager who supports and guides the interviewers. In case of any problems, enumerators are expected to immediately report them to the supervisor to seek advice and any further

instructions. The supervisor also provides enumerators with all necessary materials and instructions for work and collects and checks their work on a daily basis. Throughout the survey, there will be a range of quality control measures in place to make sure that the enumerators' work is conducted in line with procedures and that the data collected is of the highest quality. On a day-to-day basis, enumerators are monitored by the Coordination Team who assigns the interviewers and daily tasks and regularly checks their work including:

1. Interview files
2. Ensuring that some interviews are carried out by each enumerator
3. Revisiting some households to verify certain surveys

The data are checked by the Project Coordinators, Data Processing Manager and Team Leader to ensure quality and to support the enumerators in improving their performance.

Survey Solutions

This section serves as a first point of reference while starting to adapt the RIA questionnaire template to the local context in Survey Solutions. To learn more about setting-up a server, importing questionnaires and the field work capabilities of the software, it is advised to read one of the detailed articles provided by the World Bank on how to start using Survey Solutions and Frequently Asked Questions on the software, among others.

Before starting to design the questionnaire, the person scripting is strongly recommended to 1) read this guideline 2) be trained in using the Survey Solutions Designer. If the person has not used Survey Solutions before, it is strongly recommended to get accustomed with it through a dedicated training. The RIA questionnaire template itself is complex, which is al also reflected in the coding and structure in Survey Solutions and makes use of complex C# language.

KEY DEFINITION

Field Protocols are a set of rules that the research team must follow when conducting the data collection. All interviewers must follow these rules to create consistency across the data.

Box 10. Pros and Cons of Electronic Data Collection

PROS: Electronic data collection can improve data quality. In a randomized experiment designed to compare CAPI and PAPI for a consumption survey in Tanzania, Caeyers, Chalmers, and De Weerdt (2012) found that data from paper surveys contained errors that were avoided in electronic surveys. The researchers discovered that errors in the PAPI data were correlated with certain household characteristics, which can create bias in some data analysis.

Electronic data collection programs can include automated consistency checks. Certain responses can trigger warning messages so that data entry errors are minimized, and any issue is clarified with the respondent during the interview. For example, Fafchamps and others (2012) studied the benefits of consistency checks in a microenterprise survey in Ghana. They found that when consistency checks were introduced, the standard deviation of profit and sales data was lower. However, they also found that most of the times, a correction was not required: 85 percent to 97 percent of the time, respondents confirmed the original answer.

Interviews can be shorter and easier to conduct. When CAPI is used, the flow of the questionnaire can be personalized to better guide interviewers through skip patterns and to minimize mistakes and omissions in the questionnaire. In a household survey in Tanzania, CAPI interviews were, on average, 10 percent shorter than similar questionnaires collected on paper, Caeyers, Chalmers, and De Weerdt (2012) found.

Electronic data collection eliminates the need for manual reentry of data. This can reduce costs and speed up data processing.

The use of technology can bring a range of indirect benefits. For example, by using tablets or smartphones, GPS coordinates can easily be collected, or photographs can be taken. Experimental variations in the survey content can also be introduced. With some software, parts of the interview can be recorded in order to facilitate quality and monitoring checks.

CONS: The fixed costs tend to be higher for CAPI than PAPI, although the variable costs can be lower. The upfront cost of purchasing and programming electronic devices may be prohibitive for smaller impact evaluation budgets. Sufficient time is also needed up front to ensure proper programming and testing of the electronic questionnaires, which often comes after paper questionnaires have already been developed.

Specific technical expertise is needed to program electronic questionnaires and set up processes to manage the flow of data collected electronically. In developing countries with low information technology capacity, this may be difficult to find. It is also more challenging to develop software for questionnaires that are not in English or a Romance language.

Technological issues can disrupt data collection or hinder data consolidation in a secure location. Problems can arise during data collection when the electronic device has a small screen or an interface that is unfamiliar to interviewers. The risk of theft is also higher for electronic devices than paper surveys. Finally, the consolidation and synchronization of data in a secure location requires clear protocols to minimize risk of data loss. Electronic transfers of data are convenient but require a minimum level of connectivity.

Source: Gertler, Paul J., et al., Impact Evaluation in Practice, World Bank, Washington, D.C., 2016

Getting Started

It is highly advisable to **start designing the questionnaire only after finalizing the content of the questionnaire** (e.g. based on a word-file). This, of course, might be subject to time constraints in any IA and should be accounted for. Experience from the IFAD10 IA cycle showed that a person with an average understanding of Survey Solutions spent approximately two working weeks to adjust the Survey Solutions template to the project specific context, including a first batch of validation checks and skips.

Important: It is recommended to start developing new questionnaires based on the most recent version of the RIA Template. It is not advisable to use a questionnaire that was prepared for another project. The template contains important comments and steps that need to be taken into account and which can otherwise be missed.

To get access to the RIA QX Template within the Survey Solutions Designer, the Survey Solutions Focal Person of RIA can be approached via the riamailbox@ifad.org account.

It is recommended to use the questionnaire template as a starting point and ensuring that the Survey Solutions template script is tailored to the individual project, including thorough checking of the script on a rolling basis. The template itself in the Questionnaire Designer contains useful instructions that will guide in setting up the questionnaire throughout the instrument. Some useful commands and notes on Survey Solutions are provided in Section 6.4.

It is advisable to have the questionnaire translated after the training of field personnel has been carried out or at a point where further changes to the questionnaire are expected to be minimal. . In this case, the scripted questionnaire, which includes validation checks and warning messages that need to be translated, is expected to be in the final stages. To share the questionnaire for translation, it needs to be downloaded from the Survey Solutions Designer, followed by removal

> **KEY DEFINITION**
>
> A PILOT TEST is a small-scale data collection conducted to test the reliability of a data collection tool.

of duplicates (in Stata, Excel or other software) to facilitate the work of the translators. An example of instructions to be provided to translators and verification of translation is provided in section 6.4.
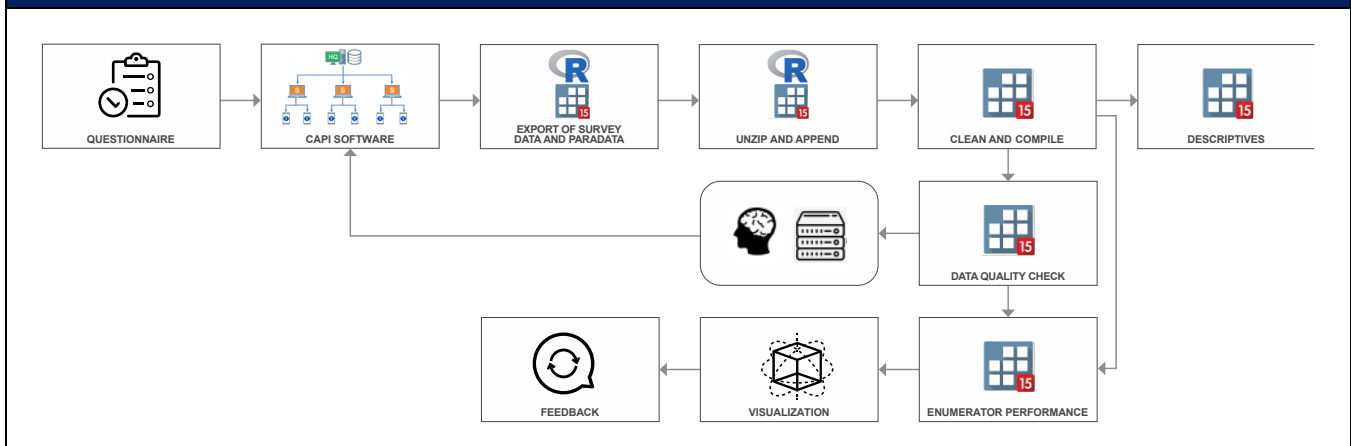
Data Quality Check System

The data collection process must be closely supervised to ensure that high quality data is captured. Having the data structure available from the first day of data collection provides a unique opportunity to conduct important data checks and to flag issues while it is still possible to post queries to enumerators and or respondents. The whole process is described in Figure 8.

Figure 8. The Data Quality System

Any data quality system is based on two complementary pillars: The CAPI software Survey Solutions, which ensures that accurate data is entered during the interview process itself, and a system of quality control measures that checks the accuracy of the data as well as enumerator performance in the post-interview process. Data quality checks mainly comprise 1) Consistency Checks, 2) Check for Missing Answers and 3) Outlier checks.

1. **Consistency Checks:** These checks are mainly implemented in the CAPI software itself. The use of validation conditions in Survey Solutions ensures that entered values are within a reasonable range, typically determined by natural bounds or secondary data during the interview process itself. Any inconsistency such as a very high number entered, or an inconsistency based on previously reported answers is flagged by the software so that the enumerator can take necessary action. If inconsistencies are detected in the post-interview process, it is advised to implement those checks in the Survey Solutions script while data collection is ongoing.

2. **Check for Missing Answers:** Missing answers, that is a question has been enabled by the software but not answered, is not accepted if no proper explanation can be provided. In general, missing answers are already flagged within the Survey Solutions Interviewer Application through the use of a green, blue and red color scheme. It is advised to emphasize on this color scheme during the training of field personnel to ensure implementation during fieldwork. However, if missing answers are flagged in the post-interview stage, interviews may need to be rejected and enumerators may need to follow up on those unanswered questions.

3. **Outlier Checks:** Numeric data from one interview are benchmarked to the rest of the obtained sample data to identify outliers, which in turn is confirmed by the enumerator and/or respondent. It has to be noted that outliers per se are not an issue during the data collection process but simply indicate that something might be wrong with the data entered. Any outlier checks that are conducted during the analysis stage can first be implemented during data collection:

   - **Univariate**: The distribution of a specific variable can be analysed, and outliers identified using the Interquartile Range (IQR) or the Standard

Deviation (SD). Subject to the nature and distribution of each variable, an outlier check could be conducted using the absolute value of the variable or its natural logarithm.

- **Bivariate**: In case that two variables are in close relation to another, they are either checked as a univariate variable using the quotient of the two or with bivariate measures (e.g., visually using box or scatter plots).

- **Multivariate**: If the data captured contains more than one variable that potentially can explain the "outcome" of a third variable, one could use those for simple multivariate regressions. Computing either Cook's or Mahalanobis distance can, in turn, be used to identify influential observations. This is of special importance for the agricultural production functions, which often form the main part of the analysis of the impact assessment. It is advised to have a basic agricultural production function set up at the end of the first week of data collection to identify influential observations and to confirm the respective data entered by enumerators.

Making use of the extensive paradata appended to each interview (provided by the Survey Solutions software package), which describes the process of data collection, allows to compute several enumerator performance indicators that will help to identify low and high performing interviewers. Those indicators comprise, but are not limited to, the following:

- Number of interviews per day (productivity)

- Working hours: Survey Solutions records timestamps of every single click on the tablet, hence allowing to detect work on interviews at uncommon times and/ or after the completion of the interviews. This, in turn, can be used to detect fabrication of interviews or the manipulation of parts of the data if no reasonable explanation can be provided by the enumerator

- Length of interviews expressed as number of answers per interview: It is assumed that the average number of answers remains constant throughout the data collection exercise. If no other reasonable explanation can be provided (e.g., different characteristics in regions sampled at the end of a survey), a decrease in the average number of answers over time would be an indication that enumerators identify pathways to purposely reduce the length of the interviews

- Speed of enumerators expressed as answers per minute: Dividing the number of answers given in an interview by the time it took to complete the respective interview helps to identify deviating behaviour such as very slow or very fast data entry

- Number of data quality checks: A high number of data inaccuracies could indicate poor understanding of instruments or problems using the device

- Systematic patterns such as the time spent at key questions or patterns in YES/NO question

- GPS Location: Could be used to confirm that enumerators follow sampling procedures

These indicators are expected to be used to benchmark the performance of individual enumerators against the rest of the sample of enumerators.

Data Tracking

The purpose of data tracking is to follow the work of the data collection firm, monitor the progress of the data collection, ensure that the firm is adhering to the sampling strategy and record the outcomes of the field visits (number of refusals and those unavailable, etc.). A separate tracking system for each instrument (i.e., one for the household survey, one for the community survey etc.) is used.

The information collected by the tracking system is tailored to the sampling strategy that is being used. An Excel spreadsheet should be used, which the firm should either put in a shared cloud folder so that their progress can be monitored in real time or can be shared on a weekly basis. The information to be collected through the tracking system is discussed with the data collection firm and the latter is expected to produce a template for approval.

Where random sampling is being used, one would want to ensure that the sampling interval and to the instructions for replacements are being adhered to. Thus, the tracking system should collect information for each household visited (if one already has lists of households, then the Household ID should also be included), including the interval of the household, the ID of the enumerator, whether or not the household is a replacement and the date and outcome of each visit to the household. If the sampling interval is calculated in a different way for each PSU, then a separate tracking sheet should be used to record the way that the interval was calculated in each case. Where the data collection firm is provided with a set list of households to sample, one would want to ensure that they are following the agreed sampling schedule, making appropriate effort to reach each household, and following the instructions for replacements.

4. ASSESSMENT

4.1 Analysis

Once data collection is finalized, the final datasets are transmitted to RIA and the **analysis** of quantitative and qualitative data is carried out by RIA analysts.

The three classes of evaluation designs that can be identified namely, **experimental**, **quasi-experimental** and **quasi-experimental**, all share the ultimate goal of trying to understand what would have happened to a particular individual, household or community, had they never received the project intervention. The so called evaluation problem needs to be solved by measuring the impact in observational data by analyzing the **difference in a specific outcome indicator comparing a situation with and a situation without the intervention**. Some main elements are:

- Data can be cross-sectional (ex-post) or panel (ex-ante)
- Measurable without/before and with/after an intervention

- Not measurable outcome if there were NO intervention (counterfactual) due to the fact that we can never observe participant who is a non-participant at the same time

Here the challenge is related to estimating causal effects since the program placement is not random and many confounding factors may affect participation in a given intervention. How the counterfactual will be constructed is also an important decision that needs to be made.

Box 11. Experimental Designs
Experimental designs concern the randomized assignment of participants to test the effects of a specific intervention on a particular outcome. Participants are randomly assigned either to one or more groups of participants, or to a control group that receives no treatment. Such evaluations are called randomized controlled trials (RCT) because of the rigorous way in which the intervention is applied – similar to how pharmaceutical products are tested. An RCT is only useful for measuring impact under certain conditions such as when i) a large sample is available; ii) the intended impacts of the intervention can be readily agreed and measured; iii) and the RCT is planned before an intervention begins.

Quasi-experimental (also called quasi-experimental) methods are always used in the case of *ex-post* impact assessments and adopted for *ex-ante* IAs in cases of ethical, political or logistical constraints. The difference between quasi-experimental methods and experimental methods is due to the fact that randomization is used to select the control group in the latter method. Quasi-experimental methods include **regression discontinuity design**, **instrumental variables**, **difference-in-differences,** and **matching** methods. Within the matching methods, the **propensity score matching (PSM)** is the one that allows to statistically create comparable groups based on an analysis of the factors behind people's propensity to participate in the project. When conducting IAs, RIA team usually uses PSM, which simulates an experimental counterfactual by constructing a control group as similar as possible to the treatment group. Once a reasonable counterfactual is created, the impacts are estimated using a set of model specifications to ensure robustness of results. These include the Nearest Neighbour Matching (NNM), Regression Adjustments

RELEVANT READINGS

For Randomized Controlled Trials (RCTs)

Angrist, J. et al. (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial American Economic Journal: Applied Economics, Vol. 1, No. 1

Duflo, E. et al. (2006). Using Randomization in Development Economics Research: A Toolkit, Department of Economics, Massachusetts Institute of Technology and Abdul Latif Jameel Poverty Action Lab, Cambridge.

(RA), Inverse Probability Weighting (IPW) and Inverse Probability Weighted Regression Adjustment (IPWRA). IPWRA is preferred given its doubly robust property ensuring consistent impact estimates even if one of the equations (treatment or the outcome) are mis-specified (Imbens and Wooldridge, 2009; Wooldridge, 2007 and 2010).

RELEVANT READINGS

Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. Econometrica, 84(2), 781-807.

Leuven, E. & Sianesi, B. (2015) PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 41-55

Imbens, G.W. & Wooldridge, J.M., (2009). Recent developments in the econometrics of program evaluation. Journal of Economic Literature, 47(1): 5-86.

Wooldridge, J.M. (2007). Inverse probability weighted estimation for general missing data problems. Journal of Econometrics, 141(2): 1281-1301.

Wooldridge, J.M., (2010). Econometric analysis of cross section and panel data. MIT press.

Rogers, P. J. (2009). 'Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation', in Chambers, Robert, et al., 'International Initiative for Impact Evaluation Working Paper No. 4, 3ie, New Delhi.

Hughes, K., & Hutchings, C. (2011). Can we obtain the required rigour without randomisation? Oxfam GB's non-experimental Global Performance Framework', International Initiative for Impact Evaluation Working Paper No. 13, 3ie, New Delhi.

Data Cleaning Guidelines

Data cleaning can be done following a series of standardized measures built on the basis of the experience of other IAs conducted. All DO files useful for data cleaning are available (riamailbox@ifad.org) and the do file organization map is described in Figure 9. For more details on the organization of the DO files, please refer to section 6.3.

Guidelines for Common Cases of Missing Data and Outliers

Specific strategies may be followed when dealing with **missing data and outliers**. According to the Rural Income Generating Activities (RIGA) methodology, income is measured as net aggregate income. Income-related variables are constructed based on information from multiple modules in the household-level questionnaire, with the objective of aggregating incomes from crop production, livestock production, livestock product production, wage employment, self-employment, and other sources (e.g. pension, remittances, gifts, etc.).

As income is a net aggregate measure, it is unlikely that the whole indicator is missing, which would result in zero income for the household. Instead, there may be households that have one or more components of aggregate income

missing. For example, a respondent may know that the household receives remittances (classified as income from 'other sources') but does not know the amount they receive. In this case, only a part of the 'other sources' component is missing, though income from crops, livestock, etc. is likely to be known for this household.

In a case like this, net income is still aggregated (i.e. the missing observation results in a value of 'zero' being assigned for that part of the income component), while acknowledging that the income indicator is underestimated for some households. There is one point of concern in this method: if households with missing values are systematically different between treatment and control groups, then it is not appropriate to treat to underestimate income. This would need to be checked for by using the Stata code provided in section 6.5 of the Annex to ensure that missing income components are distributed randomly across treatment and control groups.

Box 12. The Key Elements of a PSM

Ensure Representativeness: There needs to be a representative sample survey of eligible participants and non-participants in the intervention. Baseline data are preferred for calculating propensity scores. This technique can, however, also be used with endline data: the matching variables must be variables that are unaffected by the intervention.

Estimate Propensity Scores: The propensity scores are constructed using the 'participation equation', which is either a logit or probit regression with programme participation as the dependent variable (where in the programme = 1, not in the programme = 0). The characteristics deemed to affect participation should be well considered and as exhaustive as possible, but should exclude characteristics that may have been affected by the intervention. For this reason, it is best to use baseline data, where available, to estimate propensity scores.

Select a Matching Algorithm: Each member of the treatment group is then matched to one or more members of the comparison group. There are different ways of doing this such as matching each participant to their 'nearest neighbour' non-participant. A single individual in the comparison group may be matched to several different individuals in the treatment group.

Check For Balance: The characteristics of the treatment and comparison groups are compared to test for balance. Ideally, there are no significant differences in average observable characteristics between the two groups. Now that the treatment and comparison groups are similar on observable characteristics, variance in the outcome indicator between the treatment and comparison groups can be attributed to the intervention.

Estimate Programme Effects and Interpret Results: Finally, the impact estimate, either single or double difference, is calculated by firstly calculating the difference between the outcome indicator for the treatment individual and the average value for the matched comparison individuals, and secondly averaging out all of these differences.

Source: White, H., & S. Sabarwal (2014). Quasi-experimental Design and Methods, Methodological Briefs: Impact Evaluation 8, UNICEF Office of Research, Florence.

Outliers in the net aggregate income measure could exist at multiple levels: the individual question level (e.g. rental income), the component level (e.g. 'other sources') or the aggregate level (i.e., the total net income). Outliers may be treated by *winsorizing* at the most conservative level possible. In many cases, outliers exist only at the higher end of the distribution. Thus, it is suggested that the decision to *winsorize* both ends or only the higher end of the distribution is carefully considered.

Figure 9. DO-File Organization and Flow Chart
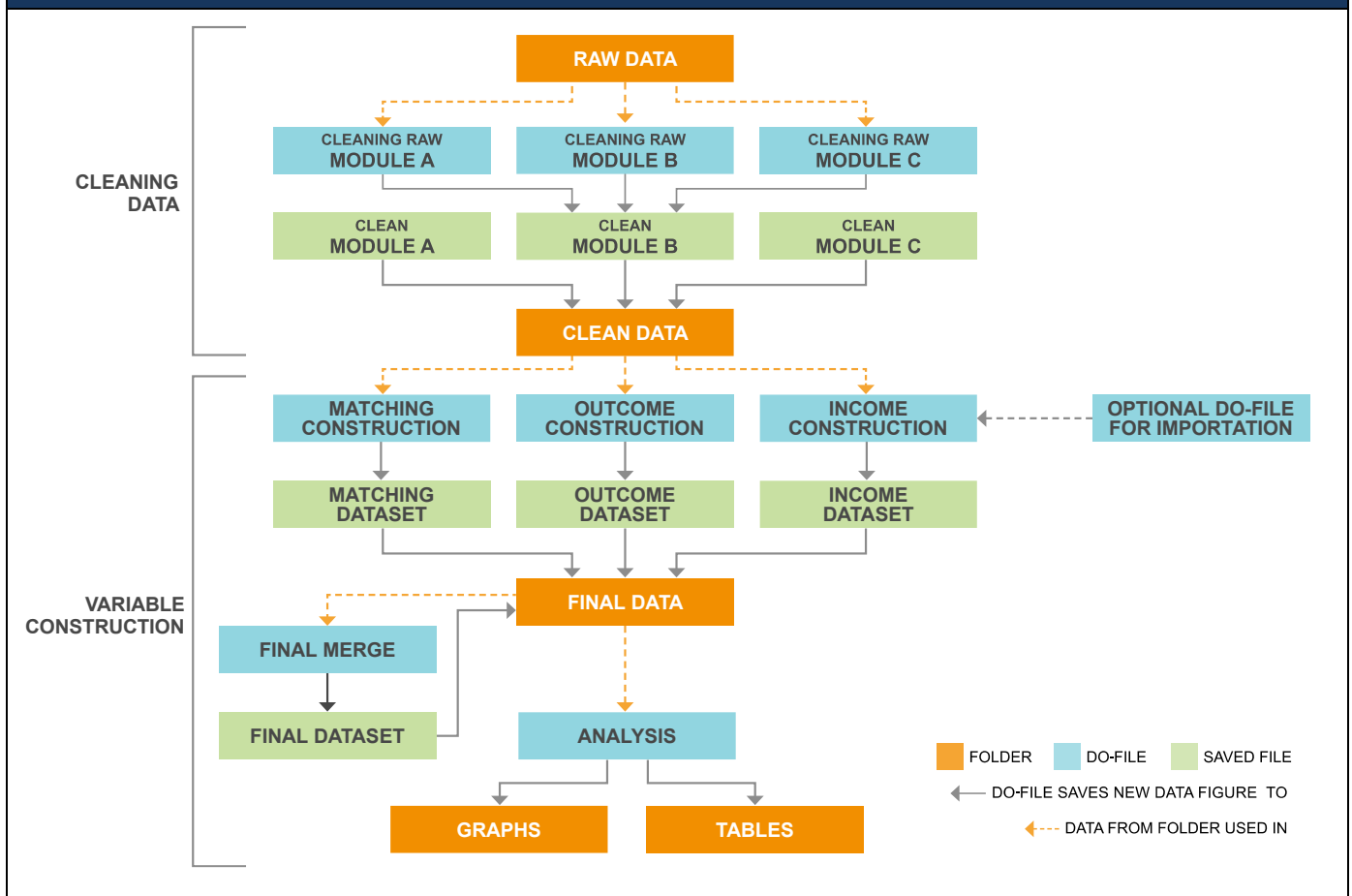
In general, outliers should only be treated at the aggregate level (ignoring those that look like outliers in individual components). However, if a component is used as its own outcome of interest, it should also be treated for outliers.

For example: A household reports income from crop production that is much higher than other observations -- that is, the data point is deemed to be an outlier. If income from crop production is not an outcome in itself in the impact assessment, then the observation is used as is as a component of income, without *winsorizing*. If the aggregate income of all 5 components is still an outlier for that household, then the observation is *winsorized* at the aggregate level.

However, if income from crop production is used as an outcome indicator in the impact assessment, it is *winsorized* before being used as a component in total net income. Stata code for *winsorizing* outliers is provided in section 6.5 of the Annex.

A common problem for the **agricultural production module** is that the destinations of a crop's harvest (collected in module C-4 of the questionnaire) do not sum up to the total amount harvested (collected in module C-3) in spite of all efforts made in Survey Solutions program with cross-validation checks. In these cases, the protocol is to trust the sum of all of the harvest destinations *over* the reported sum from module C-3. That is, total harvest is calculated by adding up all of the harvest destinations (including losses).

The price per unit for sales of crops may also be estimated incorrectly or missing. If a household indicates the sale of a portion of the crop harvested, while the price of sale is missing, the local median price can be imputed as follows:

- the price per kilogram for all observations of sales of that crop is calculated,
- the median price per kilogram from the area local to the household missing sales price is selected, and
- the local median price per kilogram is multiplied by the amount the household sold, therefore calculating the total price of sale.
- local prices for most important crops/products/wages should be collected at the community questionnaire and be used to cross-validate these values.

Deciding the geographic level to use for the estimation of local median depends on the sample size. When possible, selection of the geographic level should be an iterative process. This means that if prices are missing at the household level, price data at the village level can be used. If prices are missing at the village level, price data at the sub-district level can be used, and so on. There should be at least 10 observations of sales price at the geographic level to justify estimating the median.

For example:

Household 1 is missing price information for the sale of 100 kg of wheat. However, 12 other households from the same community also indicated they had sold wheat, and all provided the price of sale. For this community, the sales price per kg of wheat for each household that sold it is calculated and the median value is subsequently computed. Finally, the sales price of wheat sold by household 1 is imputed by multiplying this median value by the 100 kg that the household sold.

Household 2 is missing price information for the 50 kg of potatoes sold. Only 2 other households from the community sold potatoes and have price information. Eight households from the municipality sold potatoes and provided price information. Twenty households from the larger region sold potatoes and indicated the price of sale. Therefore, the price of potatoes sold by Household 2 can be imputed using the median price per kg from the

regional level only, as it is the only local level that has more than 10 observations of sales price.

If the sample size is large enough at local levels, the local median should also be stratified by the mode of sale (direct-to-market, using mediator, etc.) and by the time of sale (e.g. by month or season). Stata code to impute the price of sale for any missing sales price using a local median is provided in section 6.5 of the Annex.

Variable Construction

Assets

The components used to construct the Asset Index have values such that low values imply lower rates of ownership and vice versa. The variation and direction of individual components needs to be checked (and if not the same, recoded), rescaled and inverted if necessary, so that a value of zero implies that none of the households own a particular asset and a positive value thereafter implies higher rates of asset ownership.

RELEVANT READINGS

Austin, P.C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research, 46: 399–424.

Khandker, S. R.; Koolwal, G.B. & Samad, H.A. (2010). Handbook on Impact Evaluation: Quantitative Methods and Practices. The World Bank.

For matching, only *individual, key* assets that are recorded, are included. The end line assets are indexed and used as an outcome variable. Both recall (i.e. pre project) and current holdings of *all assets* (of durables, productive, and livestock) and housing characteristics (including access) are expected to be collected. The values and dates of purchase of each asset need not be asked. For livestock assets, big and small livestock ownership should be calculated separately (if possible).

Household Income

For what is not crop income, a reference period as the last 12 months is used (rather than the agricultural season). Income from wage employment activities for each household member are aggregated to yield household wage income.

Crop Income

Net crop income is calculated, while gross income calculations act as a robustness check. Gross income is included particularly when over 20% of net income observations have negative values.

Livestock Income

At minimum, gross livestock income questions are included with aggregate cost questions. In the case of livestock-heavy projects, net livestock income is calculated and alternative gross income is used as a check.

Business Enterprise Income

IFAD projects increasingly include household businesses/enterprises that can be processing farm products or non-farm businesses. An enterprise module collects relevant income and some cost data on business income. Gross and net profits from any household enterprise need to be carefully constructed, making sure that double counting is avoided (e.g. if a crop produced by the household is used as an input to their business, the value of crops that are processed and sold should not be in the sales from unprocessed crops component).

Descriptive and Balance Statistics

Following the impact assessment literature, the following tables and graphs are recommended for RIA-IAs. The DO-file RIA-IAs_'Descriptives.do' written using version Stata 15.1 and the descriptive statistics guidelines available can be used to produce the tables and graphs described below. Keeping in mind that not all statistics need to be included in the final IA report, there is some flexibility in the way each IA report is structured.

Statistics for IFAD impact assessment report

An impact assessment report is expected to contain informative summary statistics of variables of interest as well as statistics that illustrate the quality of balance achieved through the applied matching procedure. All tables and graphs suggested to be produced in this section are useful tools to do so, nevertheless, their inclusion in an IA report is not obligatory. They are expected to be included only if they add value to the report. Some examples from analytical work using data from a previous impact assessments are shown in this section.

Summary statistics before matching are shown in Table 6

These statistics compare treated and control units in terms of the number of observations, means, standard error, p-value of the test of difference in means and the standardized bias for variables used for matching.

RELEVANT READINGS

Contact riamailbox@ifad.org for:

RIGA methodological note

Approaches to net income construction

Table 6. Summary Statistics Before Matching				
	MEAN/SE			
	Treatment	Control	P-Value	Bias
Number of Villages	7.66	6.33	0.01***	29.25
	0.38	0.31	.	.
Number of Primary Schools	1.22	1.15	0.76	3.29
	0.14	0.18	.	.
Number of Secondary Schools	0.71	0.69	0.82	2.85
	0.06	0.07	.	.
Number of Markets	0.25	0.43	0.04**	14.22
	0.04	0.07	.	.
Distance to Primary Road	0.85	1.02	0.51	5.12
	0.16	0.20	.	.
Number of Health Facilities	0.64	0.63	0.88	3.09
	0.06	0.07	.	.
Number of Farmer Cooperatives	4.06	2.80	0.02**	25.33
	0.41	0.31	.	.
Number of Cohesive	1.73	1.72	0.98	0.07
	0.26	0.32	.	.
Population (Ln)	8.07	8.07	0.99	0.66
	0.11	0.10	.	.
Female Headed HH (Ln)	4.13	3.77	0.06*	22.12
	0.12	0.14	.	.
Arable Area	2.73	2.19	0.04**	23.90
	0.19	0.19	.	.
First Banana	0.04	0.01	0.19	13.98
	0.02	0.01	.	.
Number of Observations	106.00	102.00	.	.
Note: .01 - ***; .05 - **; .1 - *				
Point estimates are sample means. Standard errors are reported below.				
Asterisks represent level of statistical significance of t-test/chi-squared test of difference in means				

These statistics can be produced to compare the following groups if applicable:

- Beneficiaries – control
- Beneficiaries – non-eligible units
- Control – non-eligible units
- Beneficiaries – all non-beneficiaries
- Spillover group – control

They can also be produced for sub-samples, e.g. by administrative unit, crop type etc. Standardized bias allows to compare difference of sample means before and after matching and to compare the size of difference across variables.

For continuous variables:

$$B_{before}(X) = 100.\frac{\overline{X}_T - \overline{X}_C}{\sqrt{\dfrac{[V_T(X)+V_C(X)]}{2}}} \qquad B_{after}(X) = 100.\frac{\overline{X}_{TM} - \overline{X}_{CM}}{\sqrt{\dfrac{[V_T(X)+V_C(X)]}{2}}}$$

For dichotomous variables:

$$d = \frac{(\hat{p}_{treatment} - \hat{p}_{control})}{\sqrt{\dfrac{\hat{p}_{treatment}(1 - \hat{p}_{treatment}) + \hat{p}_{control}(1 - \hat{p}_{control})}{2}}}.$$

As shown in Table7 **summary statistics after matching** compare treated and control units **after matching** to illustrate the quality of balancing. Balance statistics included in the table are mean, standard deviation, standardized difference in means, reduction in bias, and test-statistic comparing distribution of p-score between treated and control.

Table 7. Summary Statistics After Matching				
	MEAN/SE			
	TREATMENT	CONTROL	P-VALUE	BIAS
Number of villages	7.35	7.18	0.78	5.06
	0.36	0.37	.	.
Number of primary schools	1.20	1.12	0.70	4.88
	0.15	0.17	.	.
Number of secondary schools	0.70	0.74	0.74	5.59
	0.06	0.08	.	.
Number of markets	0.27	0.29	0.85	3.35
	0.05	0.06	.	.
Distance to primary road	0.85	0.97	0.68	6.86
	0.17	0.20	.	.
Number of farmer cooperatives	3.72	3.90	0.79	5.10
	0.37	0.39	.	.
Population (ln)	8.06	7.89	0.61	15.24
	0.12	0.16	.	.
Female headed HH (ln)	4.12	4.00	0.60	8.36
	0.13	0.15	.	.
Arable area	2.66	2.77	0.78	5.67
	0.20	0.22	.	.
First Banana	0.03	0.03	0.98	0.55
	0.02	0.02	.	.
Number of observations	100.00	97.00	.	.
note: .01 - ***; .05 - **; .1 - *;				
Point estimates are sample means. Standard errors are reported below.				
Asterisks represent level of statistical significance of t-test/chi-squared test of difference in means.				

Table 6 and Table 7 provide information that is required for the aggregation of impact results and the meta-analysis. They can therefore be easily produced with the DO-file and then forwarded to the researcher in charge of the meta-analysis. To keep the IA reports as compact as possible, a single table is produced such that it includes all important summary statistics before and after matching as well as the reduction in bias to evaluate the quality of matching. Table 8 reports summary statistics before and after matching.

Table 8. Summary Statistics Before and After Matching									
	BEFORE MATCHING				AFTER MATCHING			REDUCTION	
	Treat. Mean/SE	Control Mean/SE	p-value	Bias	Treat Mean/SE	Control Mean/SE	p-value	Bias	in Bias (%)
Number of villages	7.66	6.33	0.008***	29.25	7.35	7.179	0.78	5.06	82.70
	0.38	0.31	.	.	0.36	0.365	.	.	.
Number of primary schools	1.22	1.15	0.762	3.29	1.20	1.117	0.70	4.88	-48.54
	0.14	0.18	.	.	0.15	0.167	.	.	.
Number of secondary schools	0.71	0.69	0.824	2.85	0.70	0.738	0.74	5.59	-96.36
	0.06	0.07	.	.	0.06	0.076	.	.	.
Number of markets	0.25	0.43	0.036**	14.22	0.27	0.287	0.85	3.35	76.44
	0.04	0.07	.	.	0.05	0.057	.	.	.
Number of Health facilities	0.64	0.63	0.883	3.09	0.63	0.656	0.79	3.66	-18.47
	0.06	0.07	.	.	0.06	0.069	.	.	.
Number of farmer coop	4.06	2.80	0.016**	25.33	3.72	3.898	0.79	5.10	79.88
	0.41	0.31	.	.	0.37	0.390	.	.	.
Number of cohesive	1.73	1.72	0.979	0.07	1.74	1.519	0.59	7.21	-9 633.92
	0.26	0.32	.	.	0.28	0.283	.	.	.
Population (ln)	8.07	8.07	0.987	0.66	8.06	7.887	0.61	15.24	-2 207.10
	0.11	0.10	.	.	0.12	0.165	.	.	.
Female headed HH (ln)	4.13	3.77	0.057*	22.12	4.12	4.002	0.60	8.36	62.23
	0.12	0.14	.	.	0.13	0.151	.	.	.
Arable Area	2.73	2.19	0.044**	23.90	2.66	2.770	0.78	5.67	76.28
Number of observations	106.00	102.00	.	.	100.00	97.000	.	.	.

A histogram of matched treated and control units on common support (using Stata commands of '*psgraph'* after '*psmatch2'*) shows whether there is a shortage of observations across the distribution. Note that the command '*psgraph'* allows to include observations that are off the common support by specifying the option '*support(varname)'*. The DO-file deletes all observations that are off the common support before drawing the graph so one would have to adjust it accordingly.

Figure 10. Histogram of Matched Treated and Control Units
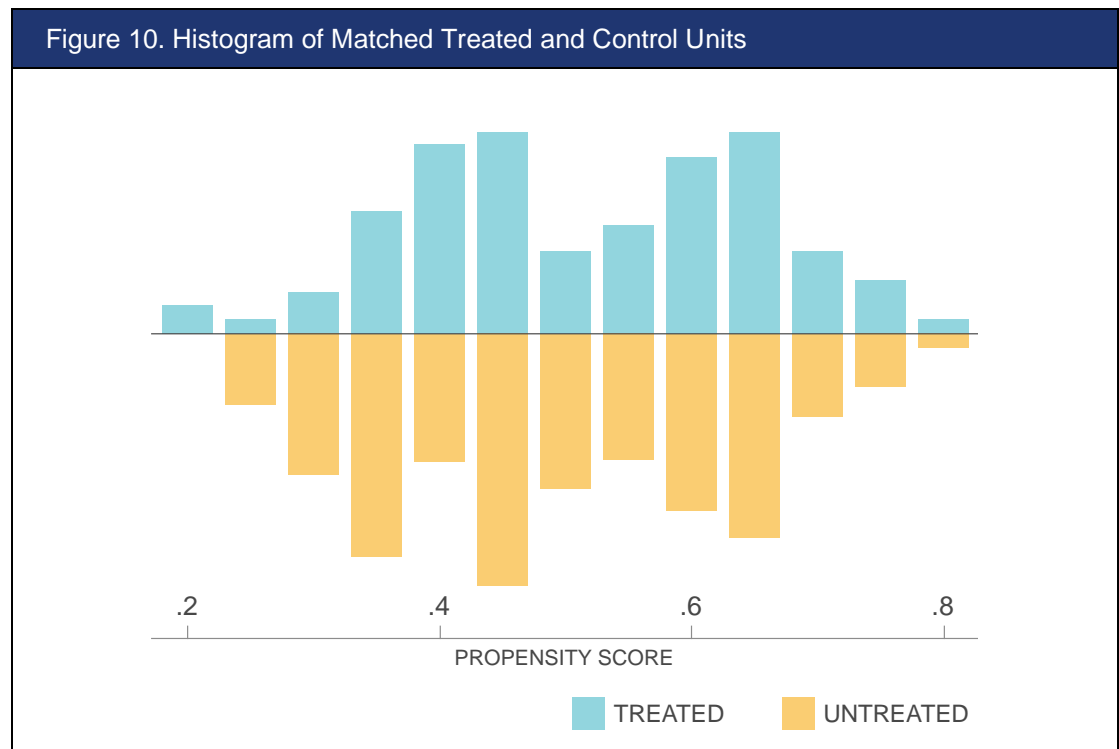
Kernel Density distribution of propensity score after matching, by treated and control (*kdensity*) is shown in Figure 11. The DO-file only generates the graph after matching but it can be easily applied to the unmatched sample as well.

Figure 11. Kernel Density Distribution of Propensity Score
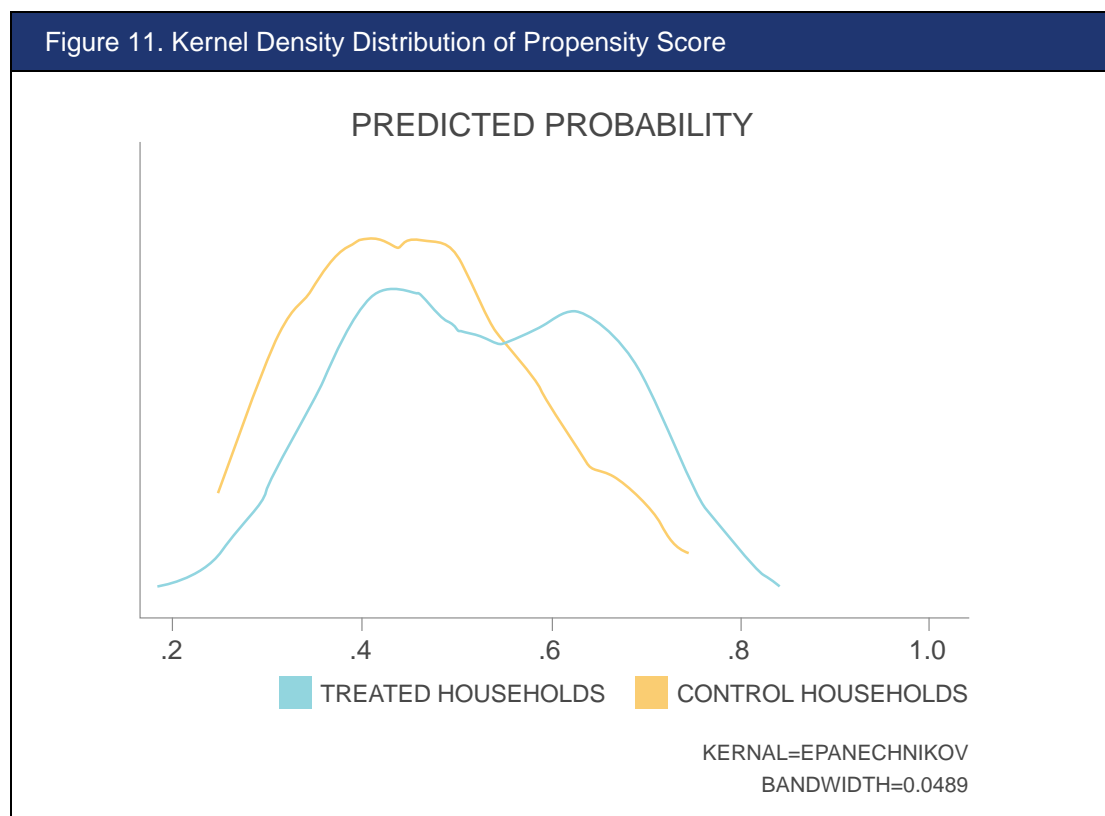
Table 9. Reduction in Bias			
	MEAN BIAS		% REDUCTION IN BIAS
	BEFORE	AFTER	
Number of Villages	29.25	5.06	82.70
Number of Primary Schools	3.29	4.88	-48.54
Number of Secondary Schools	2.85	5.59	-96.36
Number of Markets	14.22	3.35	76.44
Number of Agricultural Centres	10.80	12.33	-14.12
Distance to Primary Road	5.12	6.86	-34.08
Number of Health Facilities	3.09	3.66	-18.47
Number of Farmer Cooperatives	25.33	5.10	79.88
Number of Cohesive	0.07	7.21	-9 633.92
Population (ln)	0.66	15.24	-2 207.10
Female Headed HH (ln)	22.12	8.36	62.23
Arable Area	23.90	5.67	76.28
First Banana	13.98	0.55	96.09
First Pineapple	7.90	10.34	-30.97
First Other	4.11	5.29	-28.74
Second Pineapple	9.64	10.69	-10.91
Second Other	2.35	0.55	76.47
*Negative values indicate increase in bias after matching			

A **reduction in bias** for each covariate before and after matching can be compared in a graphic form as shown in Figure 12 using the Stata command '*pstest*'. This can also be presented in a simple table nevertheless, the graphical format is recommended.

Figure 12. Reduction in Bias
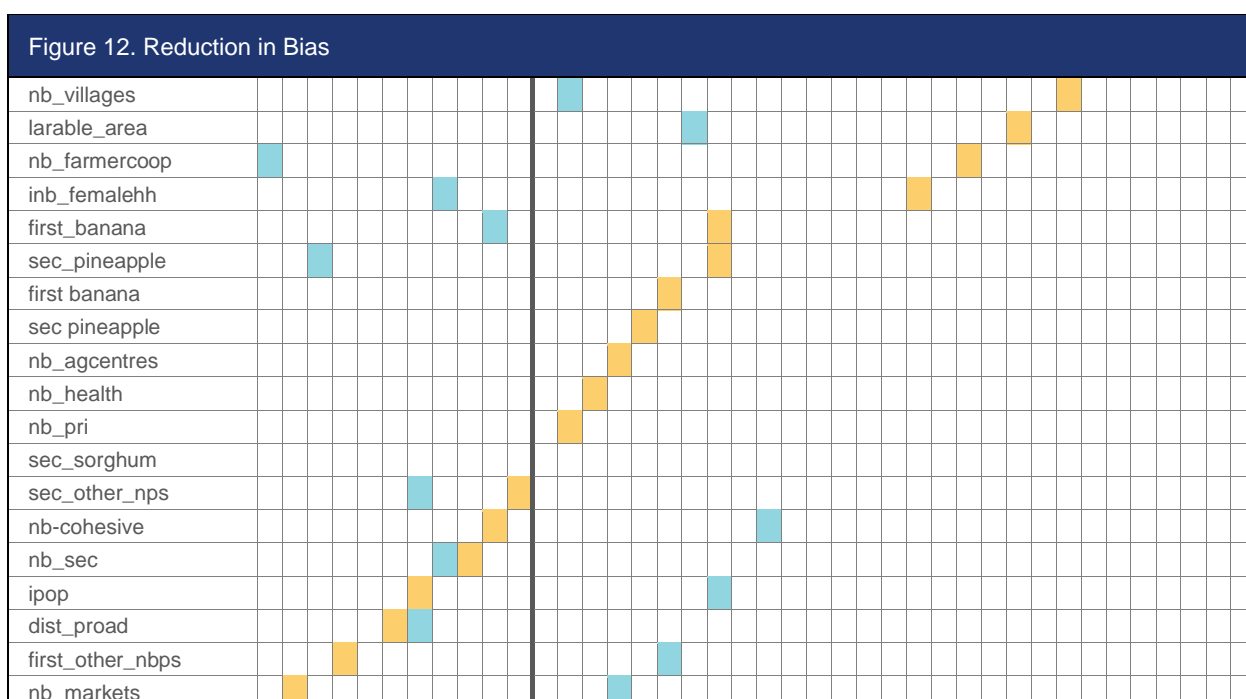
Identifying Treatment Effects

In the PSM framework[4], the impact of the project (T_i) on household i can be written as follows:

$$\delta_i = \frac{Y_{i1}}{m_i} - \frac{Y_{i0}}{m_i},$$

where δ_i is the impact of the project (or average treatment effects), Y_{i1} refers to the outcome of interest for project household i, Y_{i0} is the outcome of interest for household i in the absence of the project, and m_i is the number of observations in each cluster (in our case $m_i = 12$). Average Treatment Effects on the Treated (ATET) can be estimated using following expression:

$$ATET_{PSM} = E(\delta_i|T = 1) = E(Y_{i1} - Y_{i0}|T = 1) \qquad (1)$$

In this framework, the key identifying assumption is the conditional independence assumption (CIA) which implies that the treatment status is independent of the outcomes of interest, contingent on the observable characteristics (Rosenbaum and Rubin, 1983). Mathematically, if X_i is a vector of observable characteristics, then $T_i \perp (Y_{i0}, Y_{i1})|X_i$.

To supplement the PSM results, a regression-based analysis is also applied to consistently estimate treatment effects while controlling directly for selection into project participation based on observable characteristics. The regression method is similar to the one used in Godtland et al. (2004) to estimate the impact of farmer field schools on the returns to potato production in Peru, and in Rejesus et al. (2011) to estimate the impact of an improved irrigation technology on rice production in The Philippines.[5] Specifically, the regression specification is as follows:

RELEVANT READINGS

Rosenbaum, P.R. & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1): 41-55.

Godtland, E. M., Sadoulet, E., De Janvry, A., Murgai, R., & Ortiz, O. (2004). The impact of farmer field schools on knowledge and productivity: A study of potato farmers in the Peruvian Andes. Economic Development and Cultural Change, 53(1): 63-92.

Rejesus, R. M., Palis, F. G., Rodriguez, D. G. P., Lampayan, R. M., & Bouman, B. A. (2011). Impact of the alternate wetting and drying (AWD) water-saving irrigation technique: Evidence from rice producers in the Philippines. Food Policy, 36(2): 280-288.

Wooldridge, J.M. (2010). Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.

---

4 A presentation with more details on identifying treatment effects by employing matching can be requested from RIA (riamailbox@ifad.org).
5 See also Wooldridge (2010) for more details about this approach.

$$Y_i = \alpha + \beta T_i + \gamma \boldsymbol{X}_i + \delta(\boldsymbol{X}_i - \overline{\boldsymbol{X}})T_i + \varepsilon_i \qquad (2)$$

where Y_i is an outcome of interest, $\boldsymbol{X}_i$ is the vector of observable characteristics of household i, $\overline{\boldsymbol{X}}$ is the vector of the average of the observable characteristics of household i, and ε_i is the error term. In Equation (2), β is the ATE estimate. Replacing $\overline{\boldsymbol{X}}$ with $\overline{\boldsymbol{X}}_1$ (where $\overline{\boldsymbol{X}}_1$ is the average over treatment households only) gives us the ATET estimate.

Finally, to complement the two approaches described above, the impact of a given project by the inverse-probability-weighted (IPW) or doubly robust methods such as and the inverse-probability-weighted regression-adjustment (IPWRA) estimator (Wooldridge, 2007; Wooldridge, 2010) can also be estimated. This approach models the likelihood of being treated and estimates the impact from participating in the intervention. A major advantage of this estimation approach is that only one of the two estimation equations needs to be specified correctly, and thus has the "double-robust" property. This method follows a similar approach as the regression-based method. However, each observation in the dataset is assigned weights according to the following matrix:

$$\omega(t,x) = t + (1-t)\frac{\hat{P}(X)}{1-\hat{P}(X)},$$

where $\omega(t,x)$ is the weight applied, t represents $T_i = 1$, $\hat{P}(X)$ is the estimated propensity score, and X is a vector of covariates.

Table 10 summarizes the models used to estimate the treatment effects of the HVAP intervention in Nepal, where i denotes household, T denotes treatment indicator (1 if in the HVAP sample and 0 otherwise), Y_i denotes outcome of interest, and X_i is a vector of observable characteristics.

Table 10. Models used to estimate Treatment Effects		
METHOD	TREATMENT EFFECTS	FORMULA
Propensity Score Matching	**ATT**	$E(\boldsymbol{Y}_{i1} - \boldsymbol{Y}_{i0}\|\boldsymbol{T} = \boldsymbol{1})$
	ATE	$E(Y_{i1} - Y_{i0})$
Regression Based Method	ATT	$Y_i = \alpha + \beta T_i + \gamma \boldsymbol{X}_i + \delta(\boldsymbol{X}_i - E[X_i\|T_i = 1])T_i + \varepsilon_i$
	ATE	$Y_i = \alpha + \beta T_i + \gamma \boldsymbol{X}_i + \delta(\boldsymbol{X}_i - \overline{\boldsymbol{X}})T_i + \varepsilon_i$
Doubly Robust Method	ATT	$Y_i = \alpha + \beta T_i + \gamma \boldsymbol{X}_i + \delta(\boldsymbol{X}_i - E[X_i\|T_i = 1])T_i + \varepsilon_i$ with weights $\omega(t,x) = t + (1-t)\frac{\hat{P}(X)}{1-\hat{P}(X)}$
	ATE	$Y_i = \alpha + \beta T_i + \gamma \boldsymbol{X}_i + \delta(\boldsymbol{X}_i - \overline{\boldsymbol{X}})T_i + \varepsilon_i$ with weights $\omega(t,x) = t + (1-t)\frac{\hat{P}(X)}{1-\hat{P}(X)}$

If the outcome model is correctly specified, the RA estimator will break down more slowly than the IPW, AIPW, IPWRA, or PSM estimators as the overlap assumption begins to fail.

If the overlap assumption holds, the AIPW and IPWRA estimators have the double-robust property for some functional form combinations.

All the estimators require the same assumptions, so if each is correctly specified, they should all produce similar results. BUT just because they produce similar results does not mean that they are correctly specified!

A functional form that respects the values of the observed **outcomes is selected**. For continuous outcomes over the real line, linear is selcted; while logit, probit, or hetprobit is selcted for for binary outcomes; and poisson is selcted for for counts or nonnegative outcomes

For **binary treatments:** a choice between logit, probit, or hetprobit needs to be made. For **multivalued treatments**, only the multinomial logit is available to model the treatment probabilities.

4.2 Report

The final product is the **IA report**, which contains the results and findings on project's impacts as well as all the details of the analysis and the methodology.

Writing the Report

The IA report is expected to have the following structure:

The **executive summary** (400-500 words max) is a short summary of IA report tailored to the needs of the country team and project staff and should include:

- A short motivation to introduce the **development problem** and the IA work
- A quick summary of results
- A brief discussion of **lessons learned**

The **Introduction** (1-2 pages max) briefly describes the motivation for the impact assessment and includes:

- A quick introduction to contextualize the importance and interest in the topic and details on the relationship between the aforementioned topic and the project evaluated, while linking both to the **development problem** and the literature
- A short summary of project details and the project components that are expected to be evaluated (whole package vs. certain components);
- A brief summary of the **IA design, data and methodology** used in this impact assessment
- A roadmap to outline the structure of the rest of the IA report.

The section on **ToC and main research questions** gives a description of project interventions, targeting strategies, and coverage areas. It also includes a discussion

of the main impact assessment questions and their relevance to existing literature. The main elements included in this section are as follows**:**

- A narrative of the **ToC** explaining project logic, inputs and activities, outputs, outcomes, impacts (might differentiate between household and community levels if relevant). The focus remains on RIA team's view of the TOC, which may be different from that described in project documents. In case of a change in TOC since the drafting of the IA plan, the revised TOC is used in the report.

- A detailed description of **project components assessed**, along with specific definitions of intervention terminologies (e.g. farmer field schools, community-driven development, conservation agriculture) or any other relevant expressions. This section focuses on coverage and targeting including implications on counterfactual selection.

- The section on **project coverage and targeting** contains a description of the <u>target population</u>, targeting strategy and criteria, coverage areas while explaining all the steps involved. The key elements to be included are:

- A planned implementation schedule of project interventions (if available, especially for ex-ante design)

- A map of the intervention areas (using PMD's Factsheet format, showing the locations of the intervention areas relative to the entire country)

- A discussion of possible unintended impacts and how to address them in this impact assessment

- A discussion of possible spillover effects that may have resulted from the project (or may occur in the future for ex-ante design) and how to address them in this impact assessment

- The **main research questions** along with the main outcome and impact indicators, including references to existing research on the topic linked to the project logic.

The section on IA design: Data and methodology explains the data collection process, the overall identification strategy and methods used including the following details:

Data

- A description of the <u>overall approach</u> and the <u>identification strategy</u> to assess impacts (or planned identification strategy for ex-ante design – might include more than one contingent on the context)

- An explanation of <u>quantitative sampling design</u>, sample size, and distribution over project intervention areas

- If qualitative data are collected, a discussion of their planning, design and links to quantitative data is included to ensure credibility and representativeness

- If any other complementary data are collected (e.g. public admin. data, remote-sensing indicators, etc.), the sources and complementarity to survey data is described

Questionnaire and Impact Indicators

- The specific manner in which <u>treatment and comparison groups are created</u> and their link to project design is discussed, while emphasizing the combined use of statistics and expert consultation
- The specific manner in which <u>treatment and comparison groups are created</u> and their link to project design

Impact Estimation

- A description of the <u>statistical methodology</u> to assess impact, with a clear specification of equations estimated, level of clustering or correction of standard errors noted, impact to be estimated (ITT, ATE, ATET, LATE) clarified, methods (if distinct) for additional analyses such as spillovers and subgroup analyses stated
- If available, alternative specifications of equations to be estimated to supplement the main specifications
- A description of construction of key indicators
- A list of <u>empirical assumptions</u> necessary so as to ensure that the identification strategy adopted in the study holds
- A description of the extent to which the impact assessment design satisfies <u>internal validity and external validity</u> considerations

The section on **Profile of project area and sample** includes a short narrative about the project area and the sample used for analysis. Key elements of this section are a set of summary statistics broken down by treatment and comparison groups showing the relevant household-level and community-level characteristics and the outcome and impact indicators of interest. In the *case of ex-ante* design, both the extent to which treatment and comparison groups are comparable in terms of observed characteristics and the extent to which the desired targeting strategy is achieved, is discussed. If relevant, summary statistics are broken down by sub-groups.

The section on **Results** includes, in detail, the main findings from the impact assessment in the case of *ex-post* design and a comparison of the outcomes and impact indicators of treatment and control groups at baseline, for *ex-ante* design. In addition, the analysis of heterogeneous or sub-group effects, alternative specifications (if applicable), internal and external validity concerns, other study qualifications or limitations and important findings from the impact assessment results are discussed**.** This section is divided into sub-sections and includes**:**

> **Overall impacts of the specific project** with a description of <u>key estimates of outcome and impact indicators</u> including a clear specification of the magnitude of impacts. For matching estimators, multiple methods (at least for pooled analysis) are presented, including a triangulation of qualitative work implemented along with the quantitative questionnaire.

> **Heterogeneous impacts of the specific project** including, if relevant, a discussion of heterogeneous effects (e.g. sub-group analyses), contingent on

socio-economic and structural factors (e.g. by size of land ownership or by distance to roads or irrigation schemes). Heterogeneity analyses can focus only on preferred method/specification with a summary of other potential qualifications or limitations of the impact assessment design. In addition, important findings from impact assessment in relation to the project interventions are highlighted.

Conclusion (400-500 words max) summarizes the key lessons learned from the impact assessment, communicates the implications of the findings, provides specific, feasible and relevant recommendations to the context, and justifies cost considerations relative to project intervention (for ex-post design, if available and relevant). This specific section includes:

- A brief summary of the main impact assessment findings with reference to the development problem and the existing literature
- A discussion of key lessons learned from impact assessment findings to inform future project design, project implementation, and country strategy both for IFAD and for greater rural development policy interventions
- A context-specific, feasible and relevant set of recommendations to the implementing agency about possible ways to design and implement a project of similar type in the future

5. DISSEMINATION

5.1 Papers and Communication Output

Once the analysis is done and the IA report has been written, different outreach and dissemination tools can be produced to circulate the results of impact assessments. In particular, in addition to the IA report, a series of by-products are also produced, such as briefs, infographics, research series and journal articles. There exist different audience for results obtained through an IA : 1) country team and PMU involved in the specific program being evaluated; 2) policy makers at different levels who are expected to make use of IA results to inform funding and policy design decisions; and 3) the community of practice, including researchers and development practitioners as well as civil society participants. Each of these sets of groups are expected to benefit from outputs emanating from an IA at the end of each IA cycle, and as such, communication strategies need to be tailored to meet varied needs.

Briefs and Infographics

The RIA team, in collaboration with communication experts, produces a set of outputs in the form of policy briefs and infographics, highlighting key results from IA reports. These publications translate results of the IA report into visually appealing representations. The IA infographic template as well as the IA brief templates for each region can be requested from RIA (riamailbox@ifad.org). The impact brief has the following structure:

- Country and project name
- Short description of the project
  (max 125 words)
- The project's ToC (max 230 words)
- Project outreach and outputs (this should be a fixed text, with a list of max 9 outputs)
- Project impact (max 350 words + 2 editable figures)
- Data and methods of the estimation of project's impact
- Key impact estimates
- Lessons learned: critical section to contribute to evidence based program design

Blogs

The Strategy and Knowledge Department and the RIA division promote various content emanating from impact assessments. RIA's inputs to IFAD's expert blogs represent a means to promote RIA's technical work to a non-technical audience and to highlight the emphasis IFAD is placing, in its Knowledge Management Strategy, on the use of evidence to inform its activities and/or it's contribution to achieving the SDGs.

RIA's inputs to IFAD's expert blogs are used to engage audiences to reflect and discuss on rural development issues related to:

- evidence that has been systematically collected using rigorous and widely accepted quantitative and qualitative methods
- field experience documented in project supervision reports, mid-term reports, project completion reports or country program evaluations etc.
- raising awareness on specific topics including those linked to key events/recurrences (i.e. Women's Day, World Water Day, International Day on Remittances, etc.), new publications, results from grants etc.
- lessons and themes arising from participation in events for dissemination in a more readable manner and adapted to a broader audience than what is reported in back-to-the-office reports

In writing a blog, one must keep in mind that the **blogs' audience is smart, but not technical**. The typical reader is expected to be aware of development issues but without detailed knowledge of technical areas, research or statistical analysis.

RIA's inputs to IFAD's expert blogs is featured on several pages on ifad.org, including the expert blog gallery; the thematic pages related to impact or results; on project, country and regional websites. The Blog Submission Guidelines and the Blog Template are available along with the following checklist that provides the key elements of a well-written blog.

Guidelines for writing an expert blog are also available. A series of blogs doing mini systematic reviews of IAs and RIA's research on different types of projects (FFS, infrastructure, value chains, livestock, etc.) have been developed by the RIA team. An example of this kind of an expert blog is the one on the impact of irrigation projects published on the IFAD website. It reviews findings from a RIA impact assessments conducted in the Philippines, Ethiopia and Madagascar and synthesizes the lessons learned.

Box 14. Checklist for Blog Submission

☐ Did you use the IFAD Blog Template and Blog Submission Guidelines?

☐ Is the blog topic within the scope (evidence on achieving SDGs in rural areas)?

☐ Is the blog text 600-800 words?

☐ Is the topic clear to someone who only reads the headline?

☐ Does the lead paragraph tell who and what the story is about and why the reader should care?

☐ Would someone who is familiar with development issues but knows absolutely nothing about this topic understand this post?

☐ Is the post jargon-free?

☐ Is the post peppered with links?

☐ Have keywords already been identified to facilitate the search engine? Authors should identify at least 5-6 relevant keywords. When relevant, at least two keywords should relate to geographical coverage (one for region and the other for country e.g. #asia, #india). The rest should include project types, thematic issues, event names, etc. (e.g. #ruralfinance, #cop23, #genderequality)

☐ Did you read the blog out loud to help reveal sluggish parts of the narrative? Have you prepared a one-sentence caption for any visual media?

☐ Does your submission include a high-resolution head-shot and two sentence bio?

☐ Are there clear references and links to the sources of the evidence?

5.2 Feedback Seminar and Policy Engagement

The closing step of the cycle is a **feedback and validation seminar** to be held in the country with project's stakeholders and other key attendees. The main objective is to present and validate the main results of the impact assessment and discuss how the lessons learned can be incorporated for future design. Results of the IA are also expected to be synthesized and included in the Project Completion Report (PCR) for projects whose IA results are available before the PCR deadline. (Note that projects in the IA sample receive a 6-month extension on the PCR deadlines to facilitate this.) Once the IA report is finalized with this feedback, regional seminars are organized to share results widely with the regional divisions. Finally, participation in regional retreats and national, regional and international conferences/events is also envisaged as part of the closing phase of the cycle in order to showcase RIA's and IFAD's work.

Project Completion Reports

Results of RIA's IAs contribute to project completion reports (PCRs) with relevant information from each impact assessment. A project completion process is of key importance for identifying the ways and means to enhance the sustainability of future project interventions. It provides all stakeholders with a unique opportunity to reflect on overall project performance and generate useful lessons from implementation. The attribution analysis from the RIA IA reports complements the overall contribution analysis in the PCRs to evaluate project goals and objectives identified in the log-frame. The nuances between attribution and contribution analyses need to be carefully clarified to prevent confusion, especially in cases where the findings of each differ.

Since the adoption of IFAD's revised Evaluation Policy in 2011, the completion review process has gained even more strategic importance. IFAD's Independent Office of Evaluation (IOE) has transformed its approach from traditional project evaluations to conducting Project Completion Report Validations (PCRVs), a desk review exercise that involves a careful assessment of the quality of Project Completion Reports (PCRs). IFAD has also committed to greater efforts in measuring results and specifically impact achieved by IFAD-funded projects. Therefore, the project completion review process is being upgraded and made more rigorous in order to form the basis of independent evaluation. The Guidelines for Project Completion are being revised. An example of RIA input to PCR for the LMDPII project in Kyrgyzstan can be requested from RIA (riamailbox@ifad.org).

6. KEY REFERENCES

Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, *84*(2), 781-807.

Alkire, S., Meinzen-Dick, R., Peterman, A., Quisumbing, A. R., Seymour, G. & Vaz, A. (2013). The Women's Empowerment in Agriculture Index. World Development 52: 71–91.

Caeyers, B., Chalmers, N. & De Weerdt, J. (2012). Improving Consumption Measurement and Other Survey Data through CAPI: Evidence from a Randomized Experiment. Journal of Development Economics 98 (1): 19–33.

Fafchamps, M., McKenzie, D., Quinn, S. & Woodruff, C. (2012). Using PDA Consistency Checks to Increase the Precision of Profits and Sales Measurement in Panels. Journal of Development Economics 98 (1): 51–57.

Ferraro, P. J. & Miranda, J.J. (2017). Panel Data Designs and Estimators as Substitutes for Randomized Controlled Trials in the Evaluation of Public Programs. *JAERE,* volume 4, number 1.

Godtland, E. M., Sadoulet, E., De Janvry, A., Murgai, R., & Ortiz, O. (2004). The impact of farmer field schools on knowledge and productivity: A study of potato farmers in the Peruvian Andes. Economic Development and Cultural Change, 53(1): 63-92.

Iacus, S. M., King, G., Porro, G., & Katz, J. N. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 1-24.

Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, *106*(493), 345-361

Khandker, S.R., et al. (2010). Handbook on Impact Evaluation: Quantitative Methods and Practices, World Bank, Washington, D.C., pp. 53–103. See http://bit.ly/1d2Ve8m.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching, Political Analysis

King, G., Blackwell, M., Iacus, S., & Porro, G. (2010). cem: Coarsened exact matching in Stata.

Leuven, E. & Sianesi, B. (2015). "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing." *Statistical Software Components*.

Malapit H, Quisumbing A, Meinzen-Dick R, Seymour G, Martinez EM, Heckert J, Rubin D, Vaz A & Yount KM. (2019). Gender Agriculture Assets Project Phase 2

(GAAP2) Study Team. Development of the project-level Women's Empowerment in Agriculture Index (pro-WEAI). World Development, 122: 675-692.

Rao V. & M. Woolcock (2003). Integrating Qualitative and Quantitative Approaches in Program Evaluation", in "The Impact of Economic Policies on Poverty and Income Distribution. Evaluation Techniques and Tools, edited by Bourguignon F. and L.A. Pereira da Silva, World Band and Oxford University Press.

Rejesus, R. M., Palis, F. G., Rodriguez, D. G. P., Lampayan, R. M., & Bouman, B. A. (2011). Impact of the alternate wetting and drying (AWD) water-saving irrigation technique: Evidence from rice producers in the Philippines. Food Policy, 36(2): 280-288.

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1): 41-55.

Wooldridge, J.M. (2010). Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.

7. ANNEX: GUIDE FOR ANALYSTS

This Annex is aimed at providing more practical details for analysts to implement an impact assessment, including details on RIA Xdesk which is an important source of useful documentation. It also provides procedures for organization and naming of folders and files, together with practical recommendations for using the Survey Solutions software and Stata codes.

7.1. RIA XDesk

The Xdesk Document Environment for Sharing Knowledge is accessible through the specific link by logging in using IFAD username and password. The homepage offers a set of useful documents and information relative to each Department of the Organization. Within the Strategy and Knowledge Department (SKD), each of the three divisions, namely the Environment, Climate, Gender, and Social Inclusion (ECG), the Sustainable Production, Markets and Institutions (PMI), and the Research and Impact Assessment (RIA) allow the user to learn more about functions and activities as well as to find further details on the specific expertise and tasks. The RIA Xdesk offers an overview of the efforts made by RIA to generate knowledge by systematically assessing the impact of IFAD projects, and through research activities to support IFAD in pursuing its inclusive and sustainable rural transformation agenda. On the RIA Xdesk, detailed reports and comprehensive by-products are available and contain information on RIA's main work streams of **research**, **impact assessment** and **policy engagement**. RIA staff are responsible for ensuring all relevant documents from their IA are uploaded to Xdesk.

7.2. Folder Structure and Naming Conventions

For each file related to impact assessment activities, there exist specific guidelines and naming conventions that are followed for files and folders. This includes three

different folders within each Impact Assessment folder as described below. The Country Projects folder contains various subfolders for each of the countries IAs are being conducted for, together with the name of the project. Within each country folder, a different folder named using COUNTRY NAME_PROJECT NAME is saved. This COUNTRY NAME_PROJECT NAME folder is used at each step of the IA process and is structured as follows:

67COUNTRY NAME_PROJECT NAME

0_Project information: which includes all project background documents as well as additional documentation received from PMU or regional divisions

1_Scoping Mission: which contains documentation related to initial contact with PMU and Country Directors, but also including TORs of the scoping mission, presentations to the project team etc.

2_IA Plan: which contains the IA plan, as well as additional documentation Used To Develop The IA Plan Or Relevant For The IA Process.

3_Data Collection Firm Recruitment: which contains all the documentation for hiring the data collection firm, such as the TORs to launch the Mini tender, Work Order, etc.

4_Survey tools: which includes documents and guidelines used for quantitative and qualitative surveys, as well as questionnaires used for data collection.

5_Data Collection: which contains both household and community level data collected by RIA and secondary data. This specific folder requires to be structured further in the following sub-sections that are only uploaded to the Dropbox folder, but not on the XDesk)

> 5_1_Raw (datasets downloaded directly from Survey Solutions)
>
> 5_2_DO (DO files prepared to clean, process, and analyse datasets)
>
> 5_3_Output (processed datasets for data analysis)
>
> 5_4_Graphs (related graphs from data analysis)
>
> 5_5_Tables (related tables from data analysis)
>
> 5_6_Writing (draft IA reports, papers, etc.)
>
> 5_7_Literature (related papers, reports, and related datasets referenced)

6_IA Outputs: which contains any IA report, paper, article, brief, infographic, blog/op-ed produced during the assessment and dissemination phase, as well as feedback received by country teams and policy engagement work.

The file name also follows certain conventions. In particular, in each file name, three main components separated by "_" have to be included, namely the country name, the project name, and the file name. For country name, IFAD follows the ISO 3166-2 system. To ensure consistency, file names are assigned as follows:

- IA training
- BTOR
- IA plan
- Mini tender TOR
- HH Qx
- Comm Qx
- Qual Qx (alternatively, FGD or KII, and specify interviewee)
- IA repor

7.3. DO-File Organization

DO-files are expected to be organized using the following folder structure:

COUNTRY_NAME (Root Folder): All information for the Impact Assessment of [*COUNTRY]*

RAW_DATA: Includes all original files that the data firm sends (or that are downloaded from Survey Solutions). Contains data collected through the household survey and community survey modules.

DO: Includes all of the DO-files used in cleaning and analysis.

master.do: This is the "push-button" DO-file. It runs all of the DO-files in order.

hh_cleaning_A.do: cleans module A of the household questionnaire. A new DO-file for each module is created. Each DO-file:

Does a basic cleaning: outlier checks, fixing obvious errors, recoding 'other' options, etc.

1. Constructs the variables that are needed later for income data, matching, and other outcomes.

2. Saves the new dataset into the 'CLEAN_DATA' folder.

com_cleaning_A.do: cleans module A of the community questionnaire. A new DO-file is created for each module. Each DO-file:

- Does a basic cleaning: outlier checks, fixing obvious errors, recoding 'other' options, etc.
- Constructs the variables that are needed later for income data, matching, and other outcomes.
- Saves the new dataset into the 'CLEAN_DATA' folder.

outcome_construction.do: draws from datasets in the 'CLEAN_DATA' folder. Through creating a series of 'tempfiles', it merges relevant outcome variables (excluding the income variables) from each module into one outcome dataset entitled 'outcomes.dta' which is saved and deposited into the 'FINAL_DATA' folder.

matching_construction.do: draws from datasets in the 'CLEAN_DATA' folder. Through the creation of a series of 'tempfiles', it merges the relevant matching variables from each module into one dataset 'matching.dta' which is saved and deposited into the 'FINAL_DATA' folder.

income_construction.do: draws from the 'CLEAN_DATA' folder and merges individual modules containing information for the construction of income. It constructs income according to the RIGA method into one dataset 'income.dta' which is saved and deposited into the 'FINAL_DATA' folder.

Note: if needed, this DO-file can draw on separate DO-files for each income source (crop, livestock, wages, business, other) where price imputations, trimming, validity checks etc. are conducted. This will keep the 'income_construction.do' file from becoming too long.

final_merge.do: draws from the 'FINAL_DATA' folder and merges 'matching.dta', 'income.dta' and 'outcomes.dta'. This will create the dataset used for the final analysis, so it should include all variables needed. It saves the dataset into the 'FINAL_DATA' folder.

analysis.do: conducts all analysis, from the pre-matching descriptive statistics to the final tables. It exports graphs into the 'GRAPHS' folder and tables into the 'TABLES' folder. It also contains a log command which creates a '.log' file for each day worked on the analysis (and saves into the 'LOG' folder).

CLEAN_DATA: Cleaning DO-files deposit saved, clean data in this folder. It includes the clean modules of the household survey and community survey.

hh_A_clean.dta   **hh_B_clean.dta**

com_A_clean.dta   **com_B_clean.dta**

7.4. Notes on Survey Solutions Software

Special Survey Solutions Commands/Code-Lines

- Overall, one should be aware of the basic syntax after reading the above-mentioned guidelines. E.g. for inserting a skip it is needed to specify the "enabling condition" of a question. By writing age>15, assuming "age" contains the question of age of a respondent.

- There are two sets of syntax that may appear to be confusing and are described in detail on the links LINQ Expressions and Conditional Expressions

- **LINQ Expressions:** As described in the guidelines, the LINQ expressions are useful to search through rosters, including searching for an enabling condition or a validation check. As an example, from the RIA QX Template: parcel1. Any (p=>p. area_parcel>5). This checks the roster "parcel1". It goes through the entire roster, which one indicates by saying "p=>p" question. Please note that p as in this case is just a placeholder, one can use any letter/word. In this case, it checks if there is ANY parcel larger than 5 (standardized unit). The same expressions are used to screen for nested rosters (e.g. to lookup information at the parcel-plot-crop level).

- **Conditional Expressions:** Those are quite frequently used expressions that make facilitate validation. The basic syntax is " a ? b : c " which runs the logic: Is " a" valid/true ? If yes, do/take "b". If no, do/take "c". It often helps to compute variables or do validation checks.

- **Validation Checks in the Script:** It is highly recommended to make extensive use of validation checks within the Survey Solutions script. E.g. as a rule of thumb, every numeric question should have two validation conditions: One lower and one upper bound. However, conservative bounds are advised to be used!

- The template has a basic, and in some cases extensive, set of validation checks included. Nevertheless, every question has to be revised and adjusted to the local context.

- **Adding/Deleting Questions:** If any question of the template is deleted, it is not necessary to recode all subsequent questions. Nevertheless, if there are questions which are deleted and used for validation checks or enabling conditions later on in the template, the software will return errors after pressing the compile button which will give the person scripting the opportunity to adjust the respective conditions.

- Further, it is of utmost importance to not use the same variable name for a new question that might be added. The Survey Solutions template forms the reference point for the coding of the questionnaire. It is expected that each variable name has the same meaning across all IAs. E.g. variable "a1_3" should always be the sex of HH member. Questions added to the template shall be coded with a suffix "_m".

Formatting

- The text in Survey Solutions can be adjusted using html-codes as described in the guides. Overall, it is best practice to not use it extensively as it becomes more time intensive for the translator but most importantly maybe confusing for the enumerator.

- There is a possibility to highlight things consistently in the questionnaire, e.g. the name of the parcels or of roster items in general. It is advised to wait until the stage of uploading a final translation file to do so. Before uploading, it is recommended to manually Find And Replace the respective roster item (e.g. %parcel1%) with the formatting included (e.g. <u>%parcel1%</u>) in Excel or Stata. This will only display the formatting in the translation. One would need to do the Find and Replace in the Designer itself if translation is not used or the language is mainly used in the questionnaire designer.

Lookup Tables

A detailed description of lookup tables can be found here. Within the RIA QX template, so far there are 3 look up tables:

- **AreaConv** – Helps to standardize questions that refer to area (e.g. parcel and plot area). It is of utmost importance that local unit measurements are obtained before the start of the data collection and its respective conversion factor. Do NOT provide a "Other, specify" Option as a Unit. The current lookup table converts all units to HA. This can be adjusted and should be based on the most common local unit that most enumerators can interpret.
  - **WeightConv:** Same as above but for weight measurements. Again, try to get the local units before the data collection. In the template, the conversion table contains Liter/Ml Conversions, which is used for fertilizer/pesticides input.
- **CropConv**: Used for the conversion of crop harvest based on shelled/unshelled condition. Heavy adjustments are necessary!

To adjust the tables for an IA, the respective files can be downloaded and local measurement units and the respective conversion factors can be inserted.

Macros

- A detailed description of macros can be found here. One of the first actions to design the questionnaire should be to adjust the values and conditions to the local context, while further adjustment is often required during training of field personnel. All macros come with a description.

MODULE SPECIFIC NOTES FOR THE SURVEY SOLUTIONS SCRIPT

Cover Page: Every Cover Page contains, at the minimum, the date and time of the interview, GPS coordinates, identifying questions as well as the consent question.

In the current questionnaire template, there are 3 questions which can be considered as "Identifying". A Headquarter administrator needs to assign every supervisor/enumerator one (or more) interviews. If there are identifying questions, the HQ needs to answer those questions. If they do not answer the questions, the supervisor or eventually the interviewer has to record the answers to questions before starting the interview itself.

Consent: The consent question is mandatory. If the respondent refuses to consent, all subsequent modules are blocked and the enumerator cannot proceed with the interview. To record this incident the enumerator has to click on "complete" to close the case.

The cover page script also contains a set of variables that are used throughout the questionnaire. They are expected to be adjusted to the project, e.g. reference periods, local currency units…etc.

[A] HOUSEHOLD ROSTER

The question on household head is asked before the start of roster to enable the possibility of checking household head related questions (e.g. the age and relationship of a household member to the household head).

Relationship to household head and age of household member: There are infinite possibilities of validation checks. One should add validation checks according to local contexts.

Marital status: The validation condition must be adjusted. The idea is that if an individual has been away from the household for more than X months they should be listed in the module on migrantsinstead.

[A-3] MIGRATION
A string variable could be added that lists all members listed in the section on demographics (section A-1) who lived away from the household for more than 9 months (those should have not been listed there in the first place).

[B] PARCEL AND PLOT IDENTIFICATION
This section differentiates between Parcel and Plot use. All questions starting from module B are replicated to be asked either at the parcel or plot level. This section requires major adjustments based on the project in question and has to be checked thoroughly. The instructions within the Survey Solutions designer must be read carefully.

[C-SX] AGRICULTURAL PRODUCTION
It is advised to focus on the template of "season1" provided in Survey Solutions first and to adjust it to the local context. Thereafter, as indicated in the template, one can copy the whole module and paste it into a new blank questionnaire. Once pasted, several items can be replaced as highlighted in the template and pasted it back to the main questionnaire. This is expected to be faster as not all changes need to be made for both seasons separately.

Post-harvest crop uses and input

The roster "cropdispo_sX" shows all crops that have been cultivated in the respective season. The enabling conditions have to be tested thoroughly, especially if the template was subject to significant changes.

It is of utmost importance that the crops are coded consistently across the questionnaire. Code 100-799 are used for Seasonal Crops and 800-899 are used for Perennial Crops. 790-792 are reserved for "OTHER SEASONAL CROPS 1 -3" while 890-892 are used for "OTHER PERENNIAL CROP 1-3"

In addition, it is strongly recommended to have an exhaustive list of crops before the start of data collection to avoid using the option "OTHER, SPECIFY" for a crop question. Crops are further added to the list on a rolling basis during the pilot phase, training and the first days of data collection. The enumerators must be informed about the any crops that may be added.

Inputs Roster:
All questions concerning inputs are asked at the level of the seasonal crop. That is, questions about labor and non-labor inputs such as fertilizer, pesticides or seeds are

asked for each crop during each season. The enabling condition for non-labor inputs needs to be tested thoroughly. It currently screens all parcels and plots and checks for all those parcel/plots on which the current crop was cultivated if there was any fertilizer/pesticides used. In addition, it enables the CROP SEED row. All Crops must be added with a suffix "SEED/SEEDLING".

Random / Non-Random Sequence of Questions

There are sections in the questionnaire (e.g. O1.10-O1.16) that are expected to be asked in a random order. That is, the sequence of questions differs by interview. Several adjustments are necessary to be made as described in the script.

It is important to manually translate questions which are stored in the respective macros as this text will not show up in the regular translation file!

Once a random question sequence is chosen, it is imperative to reshuffle the data after it is exported from the server. An example to do so is provided here.

Additional Instructions for Translations of Questionnaires

- **Instructions for** Translator:
  It is advisable to send clear instructions to the translator. An example can be found below: The translator is asked to translate the whole content of column "Original Text" and insert the translation in a column next to it ("Translation").

- With respect to the contents of the cells to be translated

- Formatting:
  The translator is asked to include the html-code within the translation. If one wants to have specific words highlighted (in bold, underlined or in color), one needs to insert this in html-code within the software. As an example: If a cell contains "ENUMERATOR: RECORD START DATE & TIME FOR <u>QUESTIONNAIRE</u>", this will show up in the software as " ENUMERATOR: RECORD START DATE & TIME FOR QUESTIONNAIRE?"

- References:
  Often, we will use cross-references in the software to make use of answers that have been inserted before.

As an example: What is the sex of %name%? Here we refer to a person in the household roster. For example for the household member Mary, it would ask: What is the sex of Mary?

It is very important that the translator translates this to: %name%? Therefore*, the translator must not translate the content within ANY of the % signs.* These instructions might be cumbersome for the translator to implement but are expected to yield better quality of data."

It is not necessary to send out the columns of "Entity Id", "Variable", "Type" and Index as one can merge the translated files with the most recent template once a translation file has been received.

Changes Made to the Script While Translation is Ongoing/Finished

It is recommended that the translation template files produced by Survey Solutions are processed with Stata. Stata can be used to import the most recent translation template, which can then be merged with the most recent translated file. Doing so can help identify translations that are missing either because the original text has changed or because new questions/text items have been added.

Overall, it is advised to create a new string variable within Stata (in both the template as well as translation file) that uses the original text in lower case and without spaces and characters to better identify unique sentences/words to be used as a merge identifier as well to identify duplicates which can be dropped before the next round of translation.

Furthermore, if the question coding is added (e.g. A1.1. How old are you?) in the question text within the script, it is useful to extract the coding of the questions within Stata and then send it out for translation to facilitate the translation process. The translation tags should be added after receiving the file.

If one wants to upload a file that has the translation merged, it has to be ensured that Stata keeps all columns from the Survey Solutions template (e.g. Entity ID etc.).

Check Translation File Received

The file received by the translator must be checked to see if 1) the correct references have been inserted (or if they have added %reference% which is not in the original text) and 2) the correct html-format has been inserted. This can be done using Stata as well.

The final questionnaire used at the end of a data collection must be shared with the following email riamailbox@ifad.org.

7.5. Stata commands

Checking Distribution of Missing Values Between Treatment and Control

First, missing observations need to be explored. This is done with the 'mdesc' package. 'mdesc' produces a table with the number of missing values, total number of cases, and percent missing for each variable in varlist. It works with both numeric and character variables

For example, the number of missing in a list of variables from a to z is explored (i.e. a…z). First ensure 'mdesc' package is installed and then run mdesc a-z.

This produces a summary table with the total number of observations and the number of missing observations for each variable. To examine whether there is a systematic difference between missing observations between treatment and control groups, one would use the 'rmiss2' and 'ttest' packages. Again, make sure these packages are installed. 'rmiss2()' counts the number of missing values in the varlist. It accepts both string and numeric variables. Stata's 'rmiss()' accepts only numeric variables.

First the variable that counts the number of missing observations is generated (call it say "nmiss") using 'rmiss2()', and then the t-test using the variable that defines treatment and control (let's call this variable 'treatment' in the dataset) is carried out. The code for this exercise is provided below:

```
egen nmis = rmiss2(a-z)
ttest nmis, by(treatment)
```

Winsorizing Outliers
This can be carried out with the 'winsor' package (again make sure this package is installed). For example, to winsorize the variables a, b, and c and generate new variables for the winsorized variables, the code is the following:

```
foreach var of varlist a b c {
winsor `var' , gen(new_`var') p(0.01)
}
```

Note p(0.01) is just an example, indicating this winsorization is done at the 1st and 99th percentiles. So p(0.01) can be changed accordingly.

To winsorize observations only at the top, the option of 'highonly' needs to be added after *p(0.01), as follows:

```
foreach var of varlist a b c {
winsor `var' , gen(new_`var') p(0.01) highonly
}
```

If interested in bottom only, 'highonly' needs to be changed to 'lowonly'

```
foreach var of varlist a b c {
winsor `var' , gen(new_`var') p(0.01) lowonly
}
```

Imputing the Local Median
Imputing missing price per kg with the median by village/community/District/ municipality, province … etc. The basic goal is to generate the median price of kg crop for a locality and replace any house that sold this crop but with missing price information with this median. For instance, replacing the missing sales price of a household located in a village with the median price for all the households with nonmissing prices in the same village can be done as follows:

```
egen median_price = median(price), by(village)
```

And then the missing is replaced with the newly created median price
```
replace price = new_price if price == .
```

A Practical Example

The dataset starts with variables for each crop type, indicating the price per kg that the household received for the crop IF they sold it. So price_crop_2, for example, is only nonmissing if the household sold any of Crop 2. It will vary from household to household.

Variables for the median price of each crop within municipality, province and department are created. Then an imputed variable for the crop price is generated. If the number of observations within a region-level is over 10, the price *of the crop at that level is imputed. If none are over 10, we stop at the department level.

foreach i in 2 3 4 5 101 102 103 104 106 107 108 201 202 203 204 205 301 302 303 304 305 306 307 308 402 404 501 502 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 999
{

Each of the numbers above is a code for a specific crop as used in the survey. The median price for the levels of geographic locations is generated.

egen price_crop_`i'_mm = median(price_crop_`i'), by(A03) /* median by Municipality */

egen price_crop_`i'_mp = median(price_crop_`i'), by(A02) /* median by province */

egen price_crop_`i'_md = median(price_crop_`i'), by(A01) /* median by department */

The median crop price as equal to the median by department is generated.

gen price_crop_`i'_imp = price_crop_`i'_md

}

where A03 is the municipality variable, A02 is the province variable, and A01 is the department variable. Every crop price at the department level is imputed, and then replaced with medians from a smaller geographic region (more localized) if the observations are found to be sufficient.