European Evaluation Network for Rural Development

Funded by the European Commission

CAPTURING THE SUCCESS OF YOUR RDP:

GUIDELINES FOR THE *EX POST* EVALUATION OF 2007-2013 RDPS

JUNE 2014

Manuscript June 2014

The European Evaluation Network for Rural Development (abbreviated to “Evaluation Expert Network”) operates under the responsibility of the European Commission’s Directorate-General for Agriculture and Rural Development, Unit E.4 (Evaluation of measures applicable to agriculture; studies). The overall aim of the Network is to increase the usefulness of evaluation as a tool for improving the formulation and implementation of rural development policies by helping to establish good practice and capacity building in the evaluation of Rural Development Programmes up until 2013.

Additional information about the activities of the Evaluation Expert Network and its Evaluation Helpdesk is available on the Internet through the Europa server (http://enrd.ec.europa.eu/evaluation/)

CAPTURING THE SUCCESS OF YOUR RDP: GUIDELINES FOR THE *EX POST* EVALUATION OF 2007-2013 RDPS

JUNE 2014

CONTENT

FOREWORD

Dear Reader,

Some ninety billion Euros will have been spent on rural development by the EAFRD by the end of the programming period 2007 – 2013. An amount of which every politician (not only within rural policy), any controlling institution on national or EU level and indeed every taxpayer in Europe will want to know what were the effects of this expenditure. Evaluations in general and *ex post* evaluation in particular serve to provide this knowledge. The *ex post* evaluation also serves to summarise the success of each and every Rural Development Programme and to demonstrate what has worked well in the implementation of these programmes. As rural development policy is necessarily a continuous effort, not confined to budgetary periods, these evaluation findings are indispensable inputs into the further evolvement of policy making, hence a major source of policy learning.

For these reasons *ex post* evaluation of Rural Development Programmes (RDP) is required as a compulsory step in the implementation of each programme. With these guidelines the European Evaluation Network for Rural Development is now providing a tool for RDP stakeholders, in particular Managing Authorities (MA) and evaluators to facilitate the indeed very complex task of *ex post* evaluation.

The *ex post* evaluation guidelines draw strongly on the experience with the application of the Common monitoring and evaluation framework (CMEF) so far and on successful guidance provided earlier. It also incorporates the considerable progress of evaluation methodology made by the evaluation community during the last years (e.g. the spreading of counterfactual analysis).

The intention of these guidelines is to help the evaluation stakeholders

- to save resources for the individual programme by providing ready-to-use tools,
- to produce better and more sound findings, and last but not least
- arrive at more meaningful aggregations of evaluation findings at EU level.

Hence, the *ex post* evaluation guidelines should make the *ex post* evaluation an easier and more rewarding exercise with which the MA will be able to meet their own and EU policy needs as well as the legal requirements.

The *ex post* evaluation guidelines have been drafted by a team of evaluation experts of the European Evaluation Network for Rural Development in close collaboration with the relevant services of the European Commission and the Evaluation Expert Committee for Rural Development. Experts of the Helpdesk of the European Evaluation Network for Rural Development have contributed their evaluation experience to the text, namely John Grieve, Robert Lukesch, Jerzy Michalek, Enrique Nieto, Sari Rannanpää, Angelos Sanopoulos, Bill Slee, Peter Schneidewind, Jela Tvrdonova, Hannes Wimmer. Members of the Evaluation Expert Committee acted as a sounding board to check whether successive drafts of the text were adapted to the needs of the target audience. Representatives of DG Agriculture and Rural Development, in particular Zelie Peppiette, have ensured the coherence of the guidelines with the EU policy framework. The Helpdesk of the European Evaluation Network for Rural Development coordinated and facilitated the drafting process.

Brussels, June 2014

LIST OF ABBREVIATIONS

AEM	Agro-environmental Measures
AIR	Annual Implementation Report
AWP	Annual Work Plan
AWU	Annual Work Unit
CAP	Common agricultural policy
CCI	Common Context Indicators
CEQ	Common Evaluation Question
CF	Cohesion Fund
CGE	Computable Generals Equilibrium
CLLD	Community-Led Local Development
CMEF	Common monitoring and evaluation framework
CMO	Common Organisation of the Markets
CSF	Common Strategic Framework
DG AGRI	Directorate-General for Agriculture and Rural Development
DG BUDG	Directorate-General for Budget
DG EMPL	Directorate-General for Employment Social Affairs and Inclusion
DG ENV	Directorate-General for the Environment
DG MARE	Directorate-General for Maritime Affairs and Fisheries
DG REGIO	Directorate-General for Regional Policy
EAE	*Ex ante* evaluation
EAFRD	European Agricultural Fund for Rural Development
EC	European Commission
EIP	European Innovation Partnership
EMFF	European Maritime and Fisheries Fund
ENRD	European Network for Rural Development
EP	Evaluation plan
ERDF	European Regional Development Fund
ESI Funds	European Structural and Investment Funds
ESF	European Social Fund
EU	European Union
FADN	Farm Accountancy Data Network
FBI	Farm Bird Index

FSS	Farm Structure Survey
GAEC	Good Agricultural and Environmental Condition
GPSM	Generalised Propensity Score Method
GVA	Gross Value Added
HNV	High Nature Value
IACS	Integrated Administration and Control System
ICT	Information and Communication Technology
LAG	Local action group
LDS	Local development strategy
LULUCF	Land Use, Land Use Change and Forestry
MA	Managing Authority
MC	Monitoring Committee
M&E	Monitoring and evaluation
MS	Member State
NRN	National Rural Network
PA	Paying agency
PSEQ	Programme-specific evaluation question
PPS	Purchasing Power Standard
RDP	Rural Development Programme
SEA	Strategic Environmental Assessment
SME	Small and medium enterprise
SWOT	Strengths, Weaknesses, Opportunities, Threats
TA	Technical assistance
ToR	Terms of Reference
TWG	Technical working group
WFD	Water Framework Directive

INTRODUCTION

1. WHAT IS *EX POST* EVALUATION FOR?

Ex post evaluation is a summative evaluation of a Rural Development Programme after it has been completed. It is conducted at a point where it is possible to assess impacts and the added value of the programme funding, both at EU and programme level. The impact and added value of the interventions are important means to show programme achievements with the funds spent, justify the programme budget and enhance the transparency and accountability of EU rural policy to stakeholders and taxpayers. Demonstration of the policy achievements, thus legitimising funding for rural development measures, is important at European, national and regional levels, especially when budgets are tight. The *ex post* evaluation also provides the opportunity to see whether the policy was designed and implemented appropriately to address the most relevant needs in the programme area.

The *ex post* evaluation is a policy learning tool, which enables to use the evaluation results to improve the design, quality and implementation of the policy in the future. Utilising evaluation in policy design, via drawing relevant conclusions and lessons, is an important aspect of evidence-based policy-making. Even though the *ex post* evaluation is conducted at the end of a programming period, when the new policy is already designed and partly implemented, it has an important role in bridging the old and new programming periods. Relevant information and recommendations can be used to improve the implementation of the new programme, especially for measures which continue across programming periods. In this respect the results of *ex post* evaluation of 2007-2013 can be used as a reference in the time of preparation of enhanced Annual Implementation Reports of 2017 and 2019. Furthermore, the success and failure factors identified through the *ex post* evaluation may prove to be useful in improving the new programme and its implementation. In addition, the *ex post* evaluation can also be seen as validating or finalising the baseline values that are set for the programming period 2014-2020, since the *ex ante* evaluations were conducted before information for 2013 was available.

Although self-standing the *ex post* evaluation of Rural Development Programmes is the final part of ongoing evaluation of the programming period of 2007-2013, built upon its monitoring and evaluation activities conducted throughout the programme implementation period. The ongoing evaluation is used by evaluation stakeholders to improve the quality of Rural Development Programmes and their implementation, to justify the proposals for substantive changes during the programme life cycle and as the basis to prepare mid-term and *ex post* evaluations.

Why do we need guidelines for the ex post *evaluation of 2007-2013 RDPs?*

The assessment of the mid-term evaluation (MTE) reports conducted by DG AGRI and the Evaluation Helpdesk[1] in 2011 revealed the need to enhance the quality of evaluation since frequently the MTEs did not succeed in adequately demonstrating the achievements, results and impacts of the RDP. Therefore it was suggested to revise and reduce the number of the current set of evaluation questions outlined in the CMEF in order to better define the focus of the *ex post* evaluation. In addition, the assessment showed that Member States are actively seeking the European Commission's support in carrying out the process of the *ex post* evaluation as well as in evaluation methods.

These guidelines are drafted with significant collaborative input from the Member States, to ensure a common understanding of the requirements for the *ex post* evaluation of Rural Development Programmes and to assist Managing Authorities, evaluators and other evaluation stakeholders in preparing, steering and conducting the *ex post* evaluation. The goal is to create synergies and increase efficiency by reducing workload of the MAs and evaluators.

[1]Evaluation Helpdesk (March 2012), Methodological Assessment of Mid-Term Evaluation Reports of 2007-2013 Rural Development Programmes

The guidelines aim to provide a good understanding and a comprehensive interpretation of the legal texts in practical terms. This will save time and effort of those responsible for managing and conducting the *ex post* evaluations while ensuring that all aspects and required elements are adequately addressed. It will also help ensure consistency and comparability of evaluations among Member States and regions and facilitate the *ex post* evaluation summary[2] to be conducted by Commission services in 2017.

Despite the European Commission's involvement in the development of the guidelines, this document itself remains **non-binding**. Legally binding requirements mentioned in the guidelines are highlighted.

Who are target group(s) for these guidelines?

The *ex post* guidelines for the evaluation of Rural Development Programmes 2007-2013 have been drafted for different groups of rural development stakeholders:

1) Representatives of Managing Authorities will find information on the purpose and scope of the *ex post* evaluation, including an exhaustive list of legal references, which have to be fulfilled by Member States/regions. Furthermore, they will find the guidance and practical tools that will help them to manage, coordinate and steer the *ex post* evaluation, as well as to disseminate the evaluation results.

2) For evaluators, the guidelines provide further explanations of the legal texts and rationale behind the requirements, with the aim of creating a common understanding of the task. The document also offers a revised set of Common Evaluation Questions, clarifies the role of the evaluation questions and indicators. It also proposes evaluation methods and approaches for collecting evidence for conducting the assessment of Rural Development Programmes' impacts.

3) Other evaluation stakeholders, such as monitoring committee members, paying agencies, programme beneficiaries, etc. can use the *ex post* evaluation guidelines as a source of information, when contributing to the evaluation through data/information collection, or when debating or commenting on the *ex post* evaluation report.

4) Officials within DG Agriculture and Rural Development concerned with Rural Development Programmes 2007-2013 who may find it helpful to have a reference point summarising the common understanding of the purpose and the tasks of the *ex post* evaluation.

How to use the ex post *evaluation guidelines?*

The *ex post* evaluation guidelines have been structured with a common introduction, followed by three parts, which can be used as stand-alone documents, interlinked through cross-links provided in the text. Although each part concentrates on particular aspects of the *ex post* evaluation, which are covered in depth, in order to maintain the stand-alone nature of each part, there is inevitably a certain amount of repetition, particularly in relation to the legal requirements, in order to ensure readability and comprehension for all stakeholders involved in the evaluation of Rural Development Programmes.

The contents of the different parts are described briefly below:

Introduction introduces the guidelines and explains what the *ex post* evaluation should be conducted for.

The Scope of Evaluation explains the focus and scope of the *ex post* evaluation in terms of policy objectives, issues to be analysed and legal provisions, highlighting what must be covered. The chapter also explains common and programme-specific evaluation elements such as evaluation questions and indicators.

[2]Council Regulation (EC) No 1698/2005, Art. 87 on summary of *ex post* evaluation

PART I: **Mainly for Managing Authorities**

Chapter 1 explores the process through which the *ex post* evaluation should be conducted, providing practical guidance on the key steps of the process (planning, implementing and disseminating), the role and responsibility of each evaluation stakeholder and the timing-related issues for *ex post* evaluation.

Chapter 2 focuses on the *ex post* evaluation of National Rural Network Programmes, highlighting the differences and similarities with the evaluation of NRNs planned under the RDPs, and with the overall *ex post* evaluation process of the RDPs.

PART II: **Mainly for evaluators**

This part covers the *ex post* evaluation of Rural Development Programmes in more detail: intervention logic, evaluation questions, indicators, evaluation methods and data.

Chapter 1 deals with the **intervention logic** as the starting point of the evaluation of a programme, considering its composition, coherence between objectives and expected outputs, results and impacts, relevance in relation to addressing needs, expected and unexpected side effects of intervention logic, its effectiveness and efficiency. The chapter also provides what should be assessed about the intervention logic asking important questions and how the assessment should be approached. The chapter also provides examples of intervention logic to illustrate the proposed assessment approaches.

Chapter 2 explains the purpose and role of **evaluation questions** and judgement criteria in evaluation, and their links to indicators; it proposes a revised set of common evaluation questions, provides advices for developing programme-specific evaluation questions and gives guidance on how to use and answer evaluation questions in *ex post* evaluation.

Chapter 3 explains the purpose and role of **indicators** in evaluation and describes how to develop programme-specific indicators. The chapter also provides advice on how to use indicators in the *ex post* evaluation, including linking indicators to policy objectives and evaluation questions, balancing importance and measurability, screening indicators from the point of data requirements, etc.

Chapter 4 is dedicated to evaluation **methods** to be used in the *ex post* evaluation. The chapter starts with description of various challenges in assessing the programme impacts, is about addressing these challenges via appropriate evaluation design, highlighting importance of using the counterfactual. Further it describes qualitative, quantitative and mixed approaches in evaluation and methods to be used in the proper assessment of programme impacts. The chapter also elaborates on how to choose evaluation methods in relation to rural development axes, technical assistance and national rural networks.

Chapter 5 deals with proper use of **data**, necessary for the assessment of programme results and impacts. It also debates challenges of data management and collection and describes the role of various institutions in it.

PART III: Toolbox

The toolbox contains practical tools such as draft Terms of Reference for *ex post* evaluation, a revised set of evaluation questions, an outline of an *ex post* evaluation report, quality assessment grid for evaluation reports, examples of additional indicators for the Leader axis, and other useful tools, which will facilitate for both, evaluators and Managing Authorities, the conduction and implementation of the *ex post* evaluation.

2. SCOPE OF *EX POST* EVALUATION

The *ex post* evaluation of Rural Development Programmes should be anchored in the rural development policy objectives and focus on relevance, efficiency, effectiveness and achievements, results, impacts, as well as success and failure factors. It should be conducted using the Common monitoring and evaluation framework (CMEF). The *ex post* evaluation must, as a minimum, cover the legal requirements, which are discussed in Chapter 2.3.below.

2.1. Policy objectives

The subject of the evaluation is the rural policy developed in the form of policy objectives set up at EU, national and regional level. The policy objectives are the core of the programme intervention logic. They are linked to evaluation questions and indicators in order to conduct the evaluation.

At EU level, the policy derives from Council Regulation (EC) No 1698/2005 and is structured through a hierarchy of objectives[3], rooted in the overall objectives set up by Community strategic guidelines for rural development[4], backed by more specific objectives linked to axes and sub-axes, and continuing to operational objectives at measure level as laid down in Council Regulation (EC) No 1698/2005. As a consequence of the Health Check of the Common Agricultural Policy reform, a new set of rural development priorities reflecting recent challenges for EU agriculture and rural areas, were added to the policy priorities[5].

The objectives of rural development policy set up by Community strategic guidelines for rural development in the programming period 2007-2013 are:

- Improving the competitiveness of the agricultural and forestry sector;
- Improving the environment and the countryside;
- Improving the quality of life in rural areas and encouraging diversification of the rural economy;
- Building local capacity for employment and diversification;
- Ensuring consistency in programming (maximise synergies between axes);
- Complementarity between Community instruments.

Health Check of the CAP, Council Regulation (EC) No 74/2009, Art. 16a:

"*From 1 January 2010, Member States shall provide in their Rural Development Programmes, in accordance with their specific needs, for types of operations having the following priorities as described in the Community strategic guidelines and specified further in the national strategy plans:*

a. Climate change,
b. Renewable energies,
c. Water management,
d. Biodiversity,
e. Measures accompanying restructuring of the dairy sector,
f. *Innovation linked to the priorities mentioned in points (a), (b), (c) and (d).*"

[3]CMEF, Guidance B, 2.2 Intervention logic and indicators
[4]Council Regulation (EC)No 1698/2005, Art. 9 and Council decision 2006/144/EC of 20 February 2006 on Community strategic guidelines for rural development
[5]Council Regulation (EC) No 74/2009, The Health Check of the CAP reform

At national and regional level, the EU policy objectives set up by the Community strategic guidelines and Health Check of the CAP are transferred into Rural Development Programmes, taking into consideration specificities of the respective Member States or regions. In addition, national and regional programme authorities might include additional policy objectives addressing the specific needs of the programme territory.

2.2. Focus of evaluation

The evaluation should be an assessment at programme, axis and measure level, of the programme's:

- Relevance in terms of addressing the most important needs in the programme area;
- Effectiveness and achievements towards policy objectives;
- Efficiency in terms of receiving best value for money;
- Results in terms of programme achievements within the group of direct programme beneficiaries;
- Impacts in terms of programme contributions to the change observed in the programme area; and
- Success and failure factors and lessons learned for the future policy design.

Relevance

Programme relevance is assessed by looking at the extent to which the objectives and design of the programme are consistent with (a) challenges and concerns in a particular rural sector or programming area and (b) the needs and priorities of target groups. The assessment of relevance should include analysis of whether the objectives and the design of the programme are still appropriate at the time of the evaluation, given that circumstances may have changed since the programme was started or its objectives last revised. The appraisal should consider the programme's intervention logic: objectives, measures, activities and allocated funds, and how they address the programme area's needs identified in the SWOT analysis and needs assessment.

In evaluating the relevance of a programme it is useful to consider the following questions:

- To what extent have the programme's objectives, measures, activities and resource allocation been addressing the most important needs of the programme area over the time of programme implementation?
- To what extent has the programme generated changes which address the programme area's needs?
- To what extent are the activities and outputs of the programme consistent with the attainment of its overall objectives?
- To what extent are the activities and outputs of the programme consistent with the intended results and impacts?
- To what extent have programme activities generated/provided public goods (or addressed/corrected specific market failures) in the agricultural and rural sectors, and: Are there alternative and more efficient ways in which these could be delivered?

Effectiveness and achievements

The core of the assessment of effectiveness and achievements is the analysis of the extent to which the intervention of Rural Development Programme is heading towards the expected changes in addressing most relevant needs within the programme area and whether policy objectives and priorities defined at EU[6], national and regional level have been achieved.

The following questions guide the evaluation of effectiveness of a programme and its achievements in relation to the EU, national and regional policy objectives:

- To what extent were the objectives of the programme achieved?
- What were the major factors influencing the achievement or non-achievement of the objectives?
- To what extent has the RDP contributed to the EU rural development priorities for 2007-2013?

Efficiency

When assessing the programme efficiency, the evaluation looks at the relationship between the allocated resources and achieved programme outputs and consequently results. An analysis of programme efficiency reveals whether more outputs and results could have been obtained with the same budget, or whether the same outputs and results could have been reached with a lower cost. Efficiency is studied by looking at the size of the budget and its division between axes and individual measures, as well as the examination of budget sufficiency to achieve policy objectives and contribute to the Community priorities. The efficiency of the RDP is also affected by the delivery mechanisms and implementation procedures, so these factors should also be taken into account in the *ex post* evaluation. The results achieved through individual measures, sub-axes, axes and by the entire programme, in relation to the programme implementation costs, could also be a subject of study.

In evaluating the efficiency of a programme it is useful to consider the following questions:

- To what extent were the allocated resources able to produce expected programme results and impacts?
- Could the obtained results/impacts be produced at lower costs?
- To what extent was the contribution to the Community priorities and objectives achieved with allocated funds?
- What are the factors influencing cost efficiency of programme implementation?

Results

Assessing programme results means examining the changes happening within the group of programme beneficiaries due to the programme's interventions, in relation to the area needs, programme objectives and Community priorities. The examination can be conducted at axis, sub-axis and/or measure level in relation to specific sectors/groups of beneficiaries, or territories. The assessment of results is an essential input for the assessment of programme impacts.

In evaluating the programme results it is useful to consider the following questions:

- To what extent has the programme addressed the needs of direct programme beneficiaries?
- To what extent were the programme objectives achieved for the group of direct beneficiaries?
- To what extent have the results achieved reflected the Community priorities for rural development?

[6]Community priorities as outlined in the Community strategic guidelines for rural development and the Health Check of the CAP reform

Impacts

The *ex post* evaluation assesses the RDP impacts by scrutinising the extent to which the change observed in the programme area can be attributed to the programme and to what extent the programme has addressed Community priorities for rural development, bearing in mind the influence of other intervening factors. In the assessment of RDP impacts intervening factors also include CAP Pillar I interventions.

Impacts are linked to the wider objectives of the programme and can be observed at the programme area, macro-economic or global level (increase of incomes, employment or improvement of the quality of life in the assisted programming area). Impacts produced by a programme intervention are expressed in "net terms" after subtracting effects that cannot be attributed to the intervention (e.g. confounding factors, double counting, deadweight), and by taking into account indirect effects (displacement and multipliers). However the proportionality issue should be considered when assessing impacts of very small programmes (e.g. in case of some small regional programmes the evaluation may consider the assessment of gross impacts or programme results).

Impacts might be positive or negative, primary and secondary, expected or unexpected, intended or unintended.

The impact is typically assessed for the whole programme on axes level. In some cases the assessment of programme impacts can be also focused on a group of measures or a single measure if these contribute significantly to particular Community priorities, RDP policy objectives or if they generate substantial impacts.

In evaluating the programme impacts it is useful to consider the following questions:

- To what extent can the change in the programme area be attributed to the programme?
- What are direct and indirect effects of the programme?
- What multiplier effects has the programme produced?
- What are the intervening, confounding or external factors which influenced the programme effects?

Success and failure factors, good practice

The *ex post* evaluation also offers the opportunity to provide important lessons for designing future policies and programmes through looking at factors which contributed to the success or failure of the RDP and its interventions. This way the *ex post* evaluation plays an important role in the policy cycle and can be used for shaping policy instruments and resource allocation.

The evaluation identifies success and failure factors by analysing inherent and contextual issues which might foster or weaken the effects of the intervention. Amongst the most important inherent issues are the programme delivery mechanisms, administration and management and the effectiveness and efficiency of its operation. Other inherent issues include the level of assistance to potential beneficiaries and their capacity to absorb funds, targeting of support, information dissemination, and support in building capacity of rural development beneficiaries to develop projects. Contextual issues include the overall macro-economic situation, the ability of beneficiaries to co-finance projects, and the regional socio-economic differences.

The following questions can guide the evaluation of the success and failure factors of a Rural Development Programme:

- What were the major inherent/contextual factors fostering the RDP interventions, positively influencing the achievement of the objectives?

- What are the major objective/subjective factors weakening the RDP interventions, negatively influencing the achievement objectives?
- To which extent have objective/subjective factors fostered/weakened the programme achievements of objectives and contribution to Community priorities?
- To what extent have objective/subjective factors affected programme results/impacts?

2.3. Legal requirements regarding the scope and content of *ex post* evaluation

Ex post evaluation of Rural Development Programmes is a legal requirement[7] built on ongoing evaluation.[8] The generic aim of evaluation, including the *ex post*, is three-fold: namely, to improve the quality, efficiency and effectiveness of the implementation of Rural Development Programmes.[9]

The legal texts spell out the scope of the evaluations. In general, evaluations should assess the impact of the programmes with regard to the Community strategic guidelines[10] and the rural development objectives specific to the Member States and regions concerned. Here sustainable development requirements and environmental impacts should be taken into account.[11]

More specifically, the *ex post* evaluation, similarly to the MTE, must examine[12]:

- the degree of utilisation of resources;
- the effectiveness and efficiency of the programme;
- the socio-economic impact of the programme; and
- the programme's impact on the Community priorities**,** including new challenges[13][14].

The *ex post* evaluation, similar to the MTE, should further[15]

- cover the goals of the programme[16];
- aim to draw lessons concerning rural development policy[17];
- identify the factors which contributed to the success or failure of the programme's implementation, including as regards sustainability[18]; and
- identify best practice[19].

In addition to setting out **what** the *ex post* evaluation must examine, the legal framework also specifies **how** it should be achieved, through the legal provisions establishing the CMEF. In particular, the legal framework sets out a hierarchy of indicators which must be used in the evaluation. Quantification of all the indicators included in the implementing act (Regulation 1974/2006) in line with the definitions established as part of the CMEF is therefore an essential part of the *ex post* evaluation.

[7]Council Regulation (EC) No 1698/2005, Art. 84.1 and 86.6
[8]Council Regulation (EC) No 1698/2005, Art. 86.1 and 86.5
[9]Council Regulation (EC) No 1698/2005, Art. 84.2
[10]Council decision 2006/144/EC of 20 February 2006 on Community strategic guidelines for rural development
[11]Council Regulation (EC) No 1698/2005, Art. 84.2
[12]Council Regulation (EC) No 1698/2005, Art. 86.6
[13]Council Regulation (EC)No 1698/2005, Art. 9 and Council decision 2006/144/EC of 20 February 2006 on Community strategic guidelines for rural development
[14]Council Regulation (EC) No 74/2009, The Health Check of the CAP reform
[15]Council Regulation (EC) No 1698/2005, Art. 86.6
[16]Council Regulation (EC) No 1698/2005, Art. 4.1
[17]Council Regulation (EC) No 1698/2005, Art. 86.6
[18]Council Regulation (EC) No 1698/2005, Art. 86.6
[19]Council Regulation (EC) No 1698/2005, Art. 86.6

The values for these indicators should be analysed in conjunction with the other evaluation elements to provide the required overview of the RDP and its achievements.

Consequently, the *ex post* evaluation must contain answers to common and programme-specific evaluation questions, which are derived from an assessment of the effectiveness, efficiency, and relevance of the Rural Development Programme and provide solid conclusions and recommendations in order to improve the quality and the implementation of future programmes. The *ex post* evaluation also has to include judgements on the degree to which measures and the programme as a whole met their targets and contributed to achieving the objectives set out in the national strategy as well as the Community strategy.

The legal requirements for the *ex post* evaluation cover interventions only under Council Regulation (EC) No 1698/2005. Thus, Pillar I interventions do not need to be included. It should be noted however that inclusion of Pillar I interventions is not precluded by the legislation, and may be appropriate where complementary activities are undertaken (e.g. in relation to the Fruit and Vegetables CMO, the Wine CMO or Article 68 measures).

2.4. Common and programme-specific elements of the *ex post* evaluation

The Common monitoring and evaluation framework (CMEF) establishes an evaluation system to be used in conducting the *ex post* evaluations. The evaluation system consists of common and programme-specific elements. Common elements create a basis of the monitoring and evaluation system, and enable the comparability of information across all Rural Development Programmes in the EU. Programme-specific elements complement the common elements by addressing the specificities of the RDP in question. The programme-specific elements are designed by the programming authorities in order to highlight those aspects of the RDPs which reflect the specificities of the RDP territory, are of specific interest to rural policy stakeholders or which are not covered by the common elements.

The common and programme-specific elements comprise:

- EU common intervention logic for rural development including the hierarchy of objectives for rural development[20], the rural development measures and sub-measures;
- Common Evaluation Questions (CEQs), which are part of the CMEF and relate to Community strategic priorities, Health Check objectives, the 7 impact indicators, specific issues such as technical assistance and national rural networks, RDP efficiency, Leader and measures of the four rural development axes[21]. The CEQs focus the evaluation on the effects of programme interventions towards the EU rural development objectives and facilitate the assessment of impacts;

  Following the experience of the MTE, and in line with requests from Member States, the Commission has decided to reduce the number of Common Evaluation Questions for the *ex post* evaluation of Rural Development Programmes 2007-2013. This simplification of the common component should strengthen the comparability of the evaluation reports and facilitate aggregation at EU level, as well as providing greater opportunity for Member States/regions to use programme-specific evaluation questions designed for their particular circumstances.
- Programme-specific evaluation questions (PSEQs), are formulated by the MA to focus the evaluation on programme-specific objectives and interventions. For instance, if a programme has a specific environmental focus, the MA may wish to include additional PSEQs in the *ex post* evaluation to assess environmental effects. The PSEQs can also be used to assess

[20]Council Regulation (EC) No 1698/2005, Art. 4.1

[21]Further information on the CEQs can be found in Explanatory Notes to the Common Evaluation Questions & Recommendations on Mid-term Evaluation Reporting, European Evaluation Network for Rural Development, July 2010

elements of particular interest to the MA, such as programme implementation, delivery mechanisms, national rural network, or communication strategy.

- Common indicators[22] on the financial execution (inputs), baseline, output, results, impacts, are used to measure the programme efficiency, effectiveness, results and impacts. As stated above, these form part of the legal requirements of the system and values for them must be provided in the *ex post* evaluation.
- Programme-specific indicators[23] are specific to the Rural Development Programme in question. The programme-specific indicators are designed to answer PSEQs. In addition, they can be developed to answer CEQs, in cases where the common indicators are considered insufficient.

Common and programme-specific evaluation elements are described and explained in the CMEF in order to ensure a common approach towards the evaluation of RDPs. The Guidelines for the *ex post* evaluation of 2007 – 2013 RDP complements the CMEF and the legal framework, and has a **non-binding** character[24].

Additional guidance documents on the CMEF have been published during the programming period of 2007 – 2013 on specific issues, such as assessing impacts of RDP, capturing impacts of Leader and of measure to improve the quality of life, evaluating NRN programmes, gross value added indicator, and the HNV indicator, which could also be helpful in conducting the *ex post* evaluation.[25]

[22] Council Regulation (EC) No 1698/2005, 80 and 80.1, Commission Regulation (EC) No 1974/2006, Annex VIII
[23] Council Regulation (EC) No 1698/2005, 81.5
[24] Common monitoring and evaluation framework for rural development policy 2007-2013
[25] Additional guidance documents can be found at:
http://enrd.ec.europa.eu/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.cfm

PART I: MAINLY FOR MANAGING AUTHORITIES

1. THE PROCESS OF *EX POST* EVALUATION

1.1 Who is involved?

There are a number of actors taking part in the *ex post* evaluation. Some of them, such as the Managing Authority, the paying agency, the delegated implementing bodies (e.g. local authorities, regional development bodies, and public agencies), the Monitoring Committee, the LAGs and NRNs are obliged to take part in the evaluation. The beneficiaries, and data providers (national statistical office, relevant ministries, research institutes) are also involved in the *ex post* evaluation, and of course the evaluators are responsible for conducting it.

Evaluation steering group

Evaluation steering groups (SG) are **not mandatory**. However, an evaluation steering group is one of the ways to facilitate and coordinate stakeholder consultation and manage the evaluation process. Evaluation steering group members can contribute with their skills and expertise and help ensure the availability of data, information and relevant contacts with evaluators. An engaged evaluation steering group also enables a more effective interaction with and between partners and other stakeholders. It can therefore be considered as good practice.

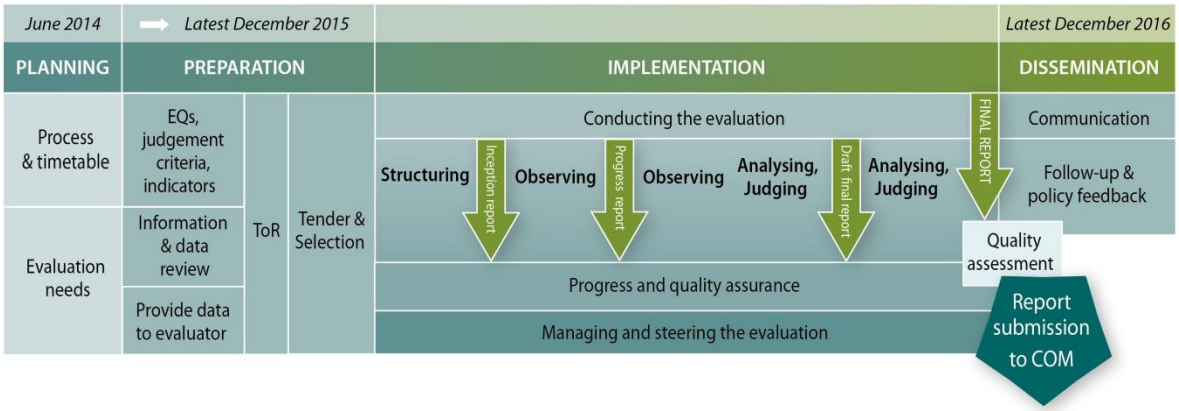
An evaluation steering group is typically convened and chaired by the MA. The composition of the group depends on the specificities of the programme (priorities, scale and delivery) and the specific tasks assigned to the group. As a minimum, the group would be expected to include representatives from the MA (representing relevant departments/units) and others involved in programme delivery, representatives from the PA, those responsible for programme design and policy-making and, if applicable, members of the evaluation unit. It may also be useful to include representatives of EU Structural Funds and experts from research institutions. There is no fixed size for the evaluation steering group. It should be large enough to cover the necessary skills and knowledge, but not as large as to hamper its effectiveness.

Some RDPs use a different steering group for each evaluation, whereas others have a single group which follows the entire ongoing evaluation process of the RDP, including the *ex post* evaluation.

1.2 Key steps

Evaluation is an ongoing process that has to be considered during the whole implementation of the programmes. Within this process, *ex post* evaluation can be seen as a separate project by itself and as such be divided into separate steps with tasks that follow each other in sequence. The main steps of the *ex post* evaluation are planning and preparation, implementation, and dissemination. Each phase contains several tasks, which will be described below together with the tasks of the main actors involved.

Figure 1 Steps, tasks and timetable related to the *ex post* evaluation



1.2.1. Planning and preparation

The MA is the main driver of the planning and preparation of the *ex post* evaluation. The better the planning and preparation are done, the clearer the implementation of the evaluation will be.

Planning the process and timetable for the evaluation

Conducting an evaluation takes a considerable amount of time. An adequate planning is vital to prevent deficiencies for instance in the collection of information to conduct the evaluation. If most information is available the evaluator would save a lot of time. Hence, the planning of the evaluation should start well in advance of the desired finalisation of the final report. Together with timetabling, it is vital to make a process plan which considers the different actors involved and which covers the entire duration of the evaluation.

The Member States must submit the *ex post* evaluation to the Commission by 31 December 2016 at the latest.[26] The Commission has a duty to complete the summary of *ex post* evaluations by 31 December 2017.[27]

The complexity, in terms of number of tasks and actors involved, necessitates timely planning in order to help anticipation of workloads and management of deadlines. Advanced planning is essential, as the whole process from evaluation planning to results dissemination can take up to three years. A lack of available data may considerably lengthen the evaluation process and it cannot be stressed enough how important it is to start planning an evaluation well in advance.

The entire *ex post* evaluation scheduling process can be managed, for instance, by using a scheduling method called retro planning, also known as backward scheduling. Retro planning means creating a schedule by starting from the last step of the process: the deadline (See example in **Part III:** Item 6). This is a useful tool in estimating the timeline for the *ex post* evaluation. Experience from previous programming periods will be helpful in determining the length of time required for each action.

- The Managing Authority is responsible for planning the process, setting up the timetable and for ensuring that the *ex post* evaluation is finished in time for the submission of the final report to the Commission.

Identification of evaluation needs

Each evaluation provides information aimed at the programme stakeholders. The first step in starting an evaluation process is to consider what kind of information needs to be produced by the evaluation. For the *ex post* evaluation of the Rural Development Programmes certain evaluation obligations are

[26] Art. 18 of the Delegated Act for Regulation (EU) No 1305/2013
[27] Art. 18 of the Delegated Act for Regulation (EU) No 1305/2013

required by the legal framework[28]. The evaluation needs are also linked to common or programme-specific policy objectives or specific policy interests (e.g. generation of employment), and to the programme delivery and administration systems.

When identifying evaluation needs, it is important to consult the rural development stakeholders as well as to review existing evaluations and studies, produced earlier on in the programming period. It is then possible to identify gaps, as well as to build on the conclusions of the previous work.

- The Managing Authority, in collaboration with other actors involved in the implementation of the RDP, reviews the results of ongoing evaluation and identifies the legal and programme-specific evaluation needs.

Review of evaluation questions and indicators

Once the kind of information the *ex post* evaluation must cover has been decided, it is necessary to think about how to obtain this information. This is done through framing appropriate evaluation questions (EQs) and linking them to indicators[29].

The evaluation questions define the focus of the evaluation and thus direct the work of the evaluator. The EQs also serve as a basis for defining the kind of information to be collected. Hence, it is vital that all evaluation questions (CEQs and PSEQs) are established and taken into account in the planning phase of an evaluation. As the number of CEQs has been reduced to a minimal set for the *ex post* evaluation, it is essential to develop PSEQs so that all identified evaluation needs are covered. It is necessary to review the CEQs and PSEQs against the specificities and objectives of the RDP so as to ensure that they will capture and reflect the effects of the RDP interventions[30]. *(See* **Part II:** Chapter 2 Evaluation questions*)*

Judgement criteria should be defined for all evaluation questions to facilitate answering them in a structured manner, enhancing transparency by making the judgement explicit, and improving the objectivity of evaluations[31]. The judgement criteria specify the aspects against which the merits of the intervention are judged. They will be used in conjunction with evidence collected (indicators and other relevant information), to answer the evaluation questions. *(See* **Part II:** Chapter 2 Evaluation questions*)*

Along with reviewing EQs and developing appropriate judgement criteria, it is important to examine the indicators needed to answer the EQs. This will give a good overview of the type and scope of the information that needs to be collected. The starting point is the set of CMEF common indicators, which must be included in the *ex post* evaluation, together with any programme-specific indicators included in the RDP. However, where some aspects are not covered by the common indicators and programme-specific indicators (e.g. new challenges introduced by the Health Check), additional programme-specific indicators may be required to collect additional information needed to answer the common and programme-specific evaluation questions. *(See* **Part II:** Chapter 3 Indicators*)*

[28] See Introduction: The scope of *ex post* evaluation
[29] See Part II: Chapter 2 on Evaluation Questions and Chapter 3 on Indicators
[30] Further information on screening evaluation questions can be found at Guidelines on the Mid-Term Evaluation of the Rural Development Programmes 2007-2013, European Evaluation Network for Rural Development, July 2009
[31] EuropeAid Guide to Evaluations

Table 1 Example: Judgement criteria linking the EQ and indicators

Evaluation Question	Judgement Criteria	Common indicator (2014-2020)	Programme-specific indicator/additional information
To what extent has the intervention contributed to preventing soil erosion and improving soil management?	Soil erosion and management has improved as a result of the intervention.	% of agricultural land under management contracts improving soil management. % of forestry land under management contracts to improve soil management.	Additional information on soil erosion of the land under management contracts. Information on soil erosion within the RDP territory. Information on links between management practices and soil erosion.

- The Managing Authority, in collaboration with other actors involved in the implementation of the RDPs (e.g. PA and intermediate bodies), reviews the common evaluation questions and common and programme-specific indicators, develops programme-specific evaluation questions and additional programme-specific indicators as necessary, and identifies appropriate judgement criteria for all evaluation questions. This task typically has to be revisited later, especially in relation to indicators, after screening the data and identifying data gaps. Where the evaluator is already selected the evaluation team can also help with this task.

It is necessary to preliminarily propose and agree on an evaluation design once evaluation questions, judgement criteria and related indicators are decided upon and once the screening of data and information, needs and potential sources and provision of data are done; It is then possible to be more precise concerning the kind of data and information sources (both qualitative and quantitative) needed to complete the evaluation, i.e. to answer the common and programme-specific evaluation questions and to provide values for all common and programme-specific indicators. The experiences on data gaps and data quality from ongoing evaluation and MTE should be taken in consideration. The evaluation design will typically be finalised only after the evaluator has been contracted.

Ensuring the availability of appropriate and high-quality data for evaluation is one of the key tasks of the Managing Authority. Its presence or absence will have a significant impact on the quality of the *ex post* evaluation exercise. The following data and information sources are typically used in RDP evaluations:

- Monitoring data, collected on RDP beneficiaries (including LAGs and NRNs) throughout the RDP life-cycle, related to programme inputs, outputs, and results. These data are collected from the monitoring system of the paying agencies or implementing bodies from application forms, payment requests, etc.;
- Disaggregated data from sector representative samples via regular surveys (e.g. Farm Accountancy Data Network (FADN), Farm Structure Survey (FSS), national surveys) or from non-beneficiaries (counterfactual analysis);
- Specific relevant data collected regularly by national institutions;
- Statistical data aggregated in suitable manner;
- Face-to-face interviews, stakeholder seminars and workshops; and
- Results of surveys and questionnaires.

Apart from data normally collected via the monitoring system or existing "in house" data (within the competence of the Ministry of Agriculture), the Managing Authority should screen all possible external data sources needed for evaluation. In this respect, it is essential to identify potential data providers. A lot of the data required for the *ex post* evaluation is collected through institutions within the competence of the ministry responsible for the implementation of the RDP. However, some data may only be accessible through external institutions[32]. In these cases, accessing the required data may prove to be challenging without inter-institutional communication, legal procedures, or funds reserved for data purchase. The agreements to purchase data should be concluded at an early point to ensure the evaluator's timely access to the data. These agreements should contain provisions on clarity of responsibility for ensuring access to data to the evaluator, as well as making sure that the evaluator can access the software or reporting tools or receive relevant extracted data from databases.

Especially when handling data related to beneficiaries or individual interviewees, it is necessary to consider the legal aspects related to data protection. In addition to Directive 95/46/EC[33] and Regulation 45/2001[34], some national rules may apply.

In connection with identification of data needs and data providers, it is often necessary to examine that the data collection methods of the data providers are suitable for providing the data required for the indicators. For the common indicators, the EU level indicator fiches should be followed.[35] All programme-specific indicators should have accompanying indicator fiches, presenting the indicator and defining the data to be collected. *(*See **Part II**, Chapter 5: Data*)*

In spite of implementing monitoring and ongoing evaluation throughout the programming period, it is possible that data for some indicators is not available. There are various reasons for this: either insufficient effort has been devoted to the issue within the ongoing evaluation system, or the existing data is not in line with the CMEF requirements or data has not been collected. To fill these data gaps two approaches are possible: either the necessary work could still be carried out by the ongoing evaluator before the *ex post* contract, or the work should be included in the ToR as a specific task for the *ex post* evaluator. In this case appropriate time and resources will need to be allocated.

- The Managing Authority is responsible for ensuring access to available data sources to the evaluator at the beginning of the *ex post* evaluation.
- The Managing Authority, in collaboration with other actors involved in the implementation of the RDPs (e.g. PA and intermediate bodies), identifies data and information needs, as well as potential data providers. As there are legal and financial aspects involved in this step, different departments of the MA/PA may have to be involved. It is also beneficial to involve data providers (e.g. institutions and research institutes) early on, especially if agreements to purchase data need to be concluded.
- The Managing Authority should ensure clear guidance on or description of the various data sources and indicators.

This step will have to be done iteratively a couple of times over the course of the evaluation planning and implementing, also with the help of the evaluator, as the chosen evaluation methods will have an effect on the data requirements.

[32] Institutions which are outside of the ministry/authority responsible for RDP
[33] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
[34] Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data.
[35] CMEF Annex 3 Indicator Guidance

Evaluation mandate

If the MA decides to use the evaluation steering group, it is advisable to draft an evaluation mandate prior to the drafting of the Terms of Reference. The evaluation mandate is a document, which gives a brief and overall description of the evaluation that will be carried out. It should specify the *scope* (what is going to be evaluated), *context* and *motives* (what are the background and motives for doing an evaluation), *responsibilities* and *timing* (how will the work be organised and in what kind of schedule), and the *objectives* (what is the expected use of evaluation). The evaluation mandate guides the preparation of the ToR, as well as the work programme of the evaluation steering group.

The evaluation mandate is typically initiated and approved by the MA, and the evaluation steering group assists the MA in the drafting process.

Terms of Reference

The *ex post* evaluation should be conducted by independent evaluator. It can be outsourced together with the ongoing evaluation in one tender procedure or it can be tendered out separately (the same evaluator as for the ongoing or a different evaluator). Each of the choices shows some advantages and disadvantages, and MAs should choose the most suitable option according to their specific context. In contracting the *ex post* evaluation, Terms of Reference (ToR) is the key document in the evaluation process. The previous tasks of the evaluation planning step described above form the basis on which the ToR can be drafted. Writing good ToR is one important thing, if not the most important one, that a Managing Authority can do to ensure a good *ex post* evaluation.

The ToR should contain a description of the context, scope and objectives of the evaluation, as well as a clear distribution of roles and responsibilities. The ToR should list and clearly describe the evaluation tasks and activities that must be conducted by the evaluator during the course of the evaluation. Evaluation questions and already available information should be specified in the ToR. Good ToR also contain the characteristics required from the evaluator, criteria for choosing the evaluator, as well as the quality assessment criteria for the final report. ToR should leave space for the evaluator to suggest evaluation methods and approaches. In a way, the drafting of the ToR is part of structuring the evaluation, a task to which the chosen evaluator will subsequently contribute as well once the contract is awarded.

The EU has not set legal requirements for the content of the ToR. However, the Guidelines on the mid-term evaluation of the 2007-2013 RDPs[36] include the recommended main elements which reflect common good practice standards.

- The Managing Authority is responsible for drafting the ToR, associated quality assessment criteria and guidelines for scoring (See **Part III:** Item 5), and ensures that the ToR reflect the realities of the RDP in question;
- The Managing Authority is charged with ensuring that sufficient human and financial resources are available for conducting the evaluation;
- The evaluation steering group (if such a body exists) supports the MA in defining and drafting the ToR.

Tendering

There are some major issues that need to be considered prior to launching the tender for the *ex post* evaluation. The main consideration for undertaking a good quality evaluation is to reserve sufficient

[36]Further information from the Guidelines on the Mid-Term Evaluation of the Rural Development Programmes 2007-2013, European Evaluation Network for Rural Development, July 2009

resources for it, which is the task of Member States[37]. The technical assistance budget of 2014-2020 RDPs can be used to finance the *ex post* of 2007-2013 RDPs. In practice the ToR should reflect the size and scope of the programme to be evaluated. It is also crucial to allocate sufficient time for the evaluation process and each evaluation stage and task. Furthermore, it is particularly important to make sure that there is enough time allocated for preparation, primary research, delays in data access and other problems, as well as interaction with the evaluation steering group *(optional).* The tender appraisal process (selection of the contractor) should take into account the resource requirements, seeking a contractor with an appropriate range of expertise who proposes to allocate sufficient time to the exercise

Before launching the tender, it is also important to set out clear rules and procedures on how the evaluator and responsible management bodies will interact. This could be done through an evaluation steering group, the evaluation unit, or a dedicated project manager.

Finally, the MA must choose the tendering procedure. This depends on the specific approach taken to engaging the evaluator (a single open call, establishment of a framework panel, direct contracting). In each case, it is fundamental to respect the relevant national procurement procedures and allocate sufficient time so that the final *ex post* evaluation is submitted to the European Commission within the deadline. Prior to the call for tender, it is important to ensure that all legal aspects related to tendering are respected.

- The Managing Authority is responsible for reserving the funds for evaluation, choosing the tendering procedure, and the management of the tendering procedure.

Checklist: Can the tendering process be launched yet?
The importance of planning the *ex post* evaluation is fundamental for the success of the evaluation. Here are a few questions which can be of help for the MA to determine whether the *ex post* evaluation planning has reached the point when the tendering process can be started.
• Have sufficient resources (human and financial) been reserved for the evaluation?
• Have the evaluation needs been thoroughly considered?
• Have the evaluation questions been looked through and PSEQs developed to cover the information needs?
• Have the indicators been scrutinised and additional programme-specific indicators developed where needed to answer all the EQs?
• Have data and information needs been examined?
• Have data gaps been identified and strategies to cover them designed?
• Have the steering and management of the *ex post* evaluation process been designed?
• Has a comprehensive ToR, reflecting the programme evaluation needs and appropriate evaluator profiles, been drafted?
• Have quality assessment criteria for the final report been drafted?
• Have the assessment criteria and process for the tenders been drafted?

[37] Articles 84(5) and 86(1) of Council Regulation (EC) 1698/2005

1.2.2. Implementing

Managing, steering and resourcing the evaluation

After the MA has contracted the evaluator, the MA should manage the evaluation process with the assistance of the evaluation steering group (if such a body exists). It is advisable to appoint an evaluation manager within the MA, who takes care of the day-to-day issues related to the *ex post* evaluation and who acts as the contact point between the evaluator and the MA. However, the MA should ensure that sufficient additional capacities and human resources are available for the *ex post* evaluation if needed.

- The Managing Authority bears the responsibility for reserving sufficient human and financial resources for the evaluation, for drawing up the contract for evaluation, as well as for the management of the evaluation from the client side.

- The evaluation steering group *(optional)* assists the MA in steering the evaluation.

Conducting the evaluation

Although the evaluation tasks are conducted by an evaluator, the Managing Authority is managing the *ex post* evaluation throughout the entire process and supports the evaluator. The process of the evaluation is structured into four distinct phases (structuring, observing, analysing, and judging[38]) and the Managing Authority has a role in each phase as follows:

The structuring phase: the Managing Authority has started the structuring work already in the preparation for the evaluation. After selecting the evaluator, the MA collaborates with its team to finalise the structure of the evaluation and makes sure that the evaluator has clear understanding of the RDP (intervention logic, including the composition of measures, modifications, and implementation), the purpose of the evaluation, evaluation questions and indicators. The Managing Authority also has to ensure that the evaluator is aware of all the evaluation tasks its team has to accomplish, and that it has access to all necessary information and data with respect to the CMEF, the planned analytical methods and tools and the evaluation design.

For its part, the evaluation team has to:

- Work in close collaboration with the Managing Authority to gain an understanding of the task at hand;
- Examine detailed intervention logics for the RDP and the individual measures;
- Define the key terms of the evaluation questions; develop judgement criteria and links to indicators for each evaluation question to be answered;
- Examine existing indicators (common and programme-specific) to be used in their assessment of the effectiveness, efficiency, and relevance of the measure and/or the programme.
- Review the data and information provided or ensured by the Managing Authority, identify data gaps and propose solutions on how to bridge them;
- Develop tools and approaches for gathering additional information and data required;
- Establish the methodology for answering the evaluation questions (CEQs and PSEQs);

The observing phase covers the collection of all available and relevant data and information. During the observing phase the Managing Authority steers the evaluation process and assists the evaluator in obtaining all available valid data additional to that provided at the start of the contract (including access to data from institutions outside of the Ministry of Agriculture). As for the evaluation team, it ensures that data and information gaps needed for evaluation, which cannot be satisfied from

[38]Further information on the four evaluation phases can be found at CMEF Guidance Note B Evaluation Guidelines

available sources (both qualitative and quantitative), are bridged using various tools and techniques (questionnaires, surveys, interviews, focus groups, case studies, etc.).

In the observing phase, the evaluator has to:

- Create the tools needed for the quantitative and qualitative analysis: interview guidelines, questionnaires, queries for extractions from databases, requests for maps, guidelines for case studies, and any other data collection instrument deemed appropriate;
- Collect data and information needed for answering each evaluation question: databases, studies, identify and contact people to be interviewed, appropriate case study areas, conduct focus groups, etc.;
- Describe the process of programme implementation over the programming period, composition of programmes, priorities and target levels, budget, financial execution, major modifications to strategy, structure and resource allocation; and
- Present quality-checked values for each of the CMEF common indicators, and programme-specific indicators.

The analysing phase ensures that the evaluator processes and synthesises all available information in a systematic way. Its team should identify the net effects of RDP interventions, and consider the effects at both beneficiary and territory level. It should use different kinds of tools and techniques to triangulate findings and thus increase confidence in the observed effects and impacts, and compare them against programme objectives and targets. In order to assess overall outcomes, the link to the baselines and trends over the programming period has to be established.

In the analysis phase, the evaluator must:

- Consider whether the establishment of appropriate typologies of measures and/or beneficiaries could reduce the complexity of empirical analysis whilst still preserving the required level of detail;
- Process and synthesise the available data and information in line with chosen evaluation methods and, where necessary, resolve remaining data gaps by modelling, extrapolations or other means; and assess against the counterfactual, as well as target levels; and
- Present estimations of the net effects of the programme.

In the final stage of the evaluation, the judging phase, the evaluator develops answers to all evaluation questions based on evidence collected with the means of indicators and draws conclusions and recommendations from the analysis, respecting the judgement criteria. The conclusions and recommendations may relate to any aspect of the RDP including its design, the effects of the programme, its axes, single measures, or the delivery mechanisms. The answers to evaluation questions must be based on sound evidence, accompanied by critical appraisal, and the limitations of the validity of the findings and the resulting judgements should be described. (See **Part II**: Chapter 4: Methods)

In the judging phase, the evaluator has to:

- Answer all evaluation questions (CEQs and PSEQs);
- Assess the effectiveness and efficiency of the programme;
- Assess measures with respect to their balance within the programme;
- Judge the degree to which the measures and the programme as a whole met their objectives;
- Identify the factors that contributed to the success or failure of the programme; and
- Draft conclusions and recommendations based on the findings.

In short, in conducting the evaluation:

- The evaluator is the main responsible for structuring, observing, analysing and judging.
- The Managing Authority and the evaluation steering group, if existing steer and backstop the evaluation**,** playing a more important role in the structuring phase and specifically in providing the initial stock of key data. In the analytical and judging phases, the Managing Authority has to look at the quality of evaluation, ensure that it respects the ToR, and that conclusions and recommendations are evidence-based.
- The PA, intermediate bodies, LAGs, NRN, beneficiaries and other data providers are vital for providing data or access to data in the observing phase.

Assuring quality and progress of the evaluation

During the *ex post* evaluation, ongoing dialogue between the evaluator, the Managing Authority and the evaluation steering group, if existing, should be ensured.

The evaluation approaches and chosen methods have implications for data collection. There should be enough time for developing and discussing them with the MA and the evaluation steering group. The evaluator should be in frequent contact (with the MA and submit agreed interim deliverables as specified in the ToR). Different reports (inception report, intermediate report(s), draft final report, and final report) and presentations agreed in the contract should be submitted by the evaluator at key stages of the evaluation process. These reports should be assessed by the evaluation steering group and the MA, and timely and relevant feedback should be given to the evaluator. This helps monitoring and improving the quality of the evaluation. Further progress of the evaluation can be checked frequently against agreed milestones.

- The evaluator is responsible for submitting agreed deliverables and taking feedback into account.
- The Managing Authority and the evaluation steering group should give meaningful and timely feedback to the evaluator.
- The evaluation steering group monitors the progress of the *ex post* evaluation. *(optional)*
- The Managing Authority should facilitate the dialogue between the evaluator and the evaluation steering group, PA, intermediate bodies, LAGs, NRN, beneficiaries and data providers.

Quality assessment

The quality of the evaluation work should be checked throughout all stages of its implementation. In addition, the final report should undergo a thorough quality assessment, preferably using assessment criteria that were included in the call for tender, including those assessing the evaluation methods applied to conduct evaluation[39]. For this purpose, it is recommended to the MA to develop quality standards for evaluation reports and/or a quality assessment grid[40]. It is also good practice, to develop and document assessment criteria. There are no common compulsory quality standards or quality assessment criteria for *ex post* evaluation reports.

Good quality criteria consider the evaluation process (relevance, timeliness and inclusiveness), normative issues (focus on independence and impartiality of the evaluator), as well as technical criteria (relevance of the evaluation, appropriate design, reliable data, sound analysis, credible findings, evidence-based answers to evaluation questions, valid conclusions, helpful recommendations, and report clarity[41]). It is good practice to employ the quality criteria with a scale of

[39] Part II: Chapter 4.5
[40] Part III: Item 4
[41] Annex 6 'Quality Assessment Form', DG Markt Guide to Evaluating Legislation, pages 87-97

rating parameters (yes/no, numerical, rating scale) and standardise the requirements for each item of the scale to ensure the transparency of the quality assessment.

A draft quality assessment grid is included as an example in **Part III**, Toolbox of this document.

- The Managing Authority and the evaluation steering group *(optional)* should assess the quality of the final *ex post* evaluation report, using an agreed quality assessment procedure.

Report submission

The evaluator should submit the final report by the time agreed in the ToR. This deadline should foresee the discussion of the final report by the evaluation steering group *(optional)* and the MC[42], as well as discussion with other national/regional bodies, before the report is submitted to the EC by the end of December 2016.[43]

- The Managing Authority is responsible for submitting the *ex post* evaluation report to the Commission in time.

1.2.3. Dissemination

Communication of evaluation findings

Evaluation results are most useful if they are communicated to the appropriate target audiences in a timely manner. Developing an appropriate communication plan for evaluation results is therefore an essential part of evaluation activity planning. Even though the dissemination step logically comes after the finalisation of the *ex post* evaluation report, it is necessary to start devising the communication plan for the *ex post* evaluation findings well before the submission of the report.

The first step in establishing a communication plan for evaluation is to identify the key potential users (the target audience: who for) and their information needs (what). The main target audiences for the *ex post* evaluation findings are typically key policy-makers and interested institutions, RDP stakeholders, other interest groups, and the general public. After identifying what kind of issue would be of interest to different target audiences, the channels of communication suitable for each audience should also be outlined (how). Depending on the target audience, different means of diffusing evaluation findings can be used (e.g. meetings, synthesis notes, memoranda, presentations, brochures, newspaper articles, press conferences, newsletters, web sites, tweets, etc.). Finally, the timing of the different means of communication (when) and the persons responsible should be decided (who). The main elements for developing a communication plan can be presented in a table format.

Table 2 Elements of communication strategy

WHO	WHO FOR	WHAT	WHEN	HOW

- The Managing Authority is responsible for developing and implementing the communication plan for *ex post* evaluation. Usually, the MA has a communications department, which handles the evaluation communications, as well as the general RDP communication;

- The National Rural Network, the LAGs and the evaluation steering group *(optional)* can assist the MA in communicating the evaluation findings.

[42] Council Regulation (EC) No 1698/2005, Art. 78(c)
[43] Art. 18 of the Delegated Act for Regulation (EU) No 1305/2013

Follow-up of evaluation findings and feedback into policy

Even though *ex post* evaluation is an assessment of the programming period that has already ended, it is recommended to consider an internal procedure through which the relevant evaluation findings feed into the policy cycle (i.e. for use in the new programming period). This is particularly the case, for any measures or interventions which are maintained in similar format, and with regard to the delivery mechanisms and the management aspects of the programmes, which often remain similar from one programming period to the other.

One way of following up evaluation findings is first to go through the *ex post* evaluation report and consider the relevant findings and recommendations for the new programming period. Reasons should be given for ignoring the implications of a recommendation for the new programme. Afterwards, these recommendations should be put on the annual work list of the MA (or other relevant bodies) with a timetable for achievement. The progress of fulfilling the recommendations should be included in the annual reporting of the institutions or bodies in question.

- The **Managing Authority** should develop a plan and process for following up the evaluation recommendations.

PART II: MAINLY FOR EVALUATORS

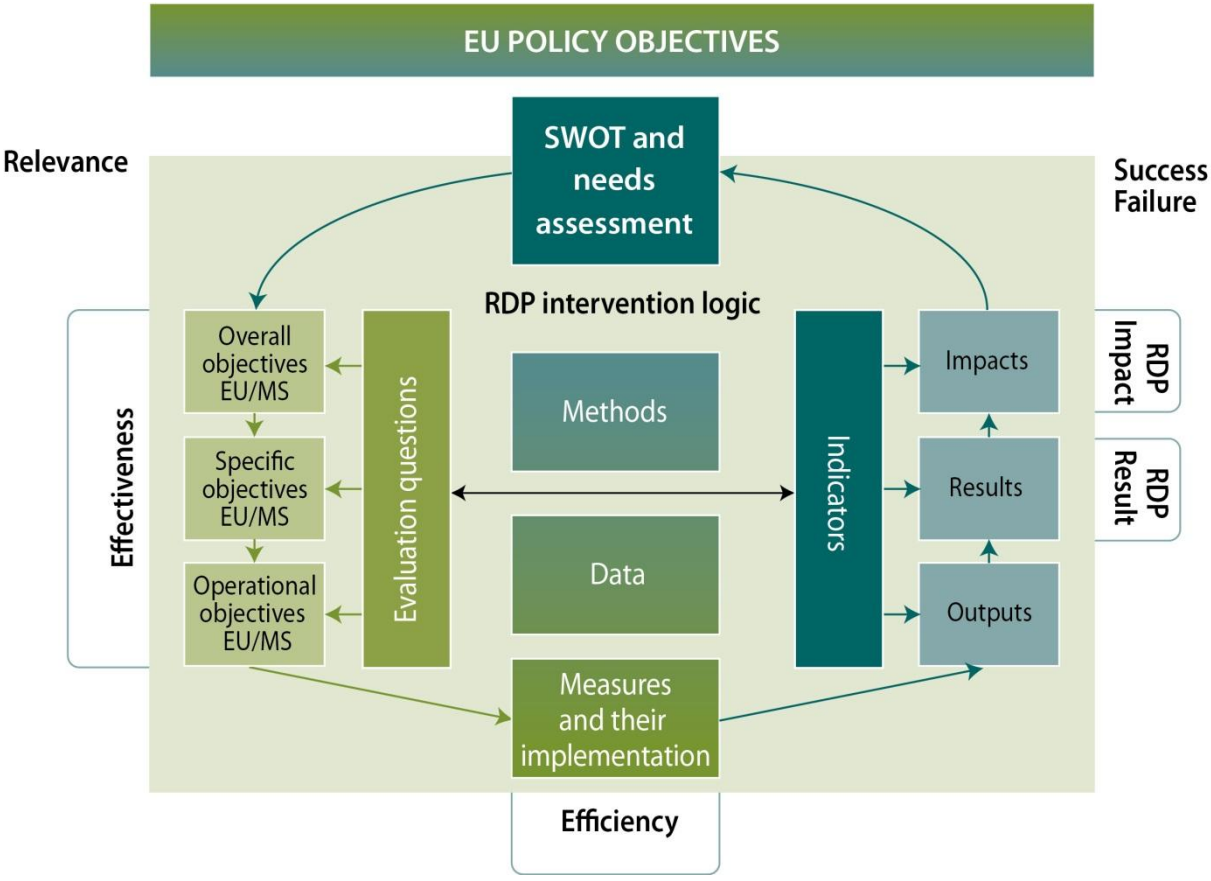
INTRODUCTION OF PART II

Part II of the *ex post* evaluation guidelines is dedicated mainly to evaluators and elaborates on the intervention logic and its relevance as the basis of the *ex post* evaluation, on evaluation questions defining the focus of evaluation and on indicators as principal means to measure efficiency, effectiveness, achievements, results, impacts and factors of success and failure of rural development interventions. **Part II** also highlights methodological challenges in the assessment of direct and indirect programme effects, separation of factors affecting evaluation results, suggests evaluation design and evaluation methods and gives advice on their selection. In addition **Part II** provides guidance on data necessary for evaluation of programme results and impacts. Chapters of **Part II** also take in consideration the specific features of evaluation of technical assistance and national rural networks.

Although **Part II** mainly targets evaluators, Managing Authorities of Rural Development Programmes can find useful information for the preparation of Terms of Reference to select the evaluator, as well as in the stage of planning, preparing and steering the *ex post* evaluation. In addition **Part II** can help Managing Authorities as well as other rural development stakeholders involved in evaluation (Monitoring Committees, evaluation steering groups) in assessing the quality of the *ex post* evaluation reports.

The overview of all important parts of the *ex post* evaluation discussed in **Part II** is presented in the Figure 2 below

Figure 2 **Parts of the *ex post* evaluation and their relation**



Source: Helpdesk of the European Evaluation Network for Rural Development

1 INTERVENTION LOGIC

The purpose of this chapter is to explain EU rural policy intervention logic in its various possible forms and to provide guidance on what should be assessed in relation to the intervention logic and how the assessment should be conducted.

1.1. Intervention logic in the EU rural policy

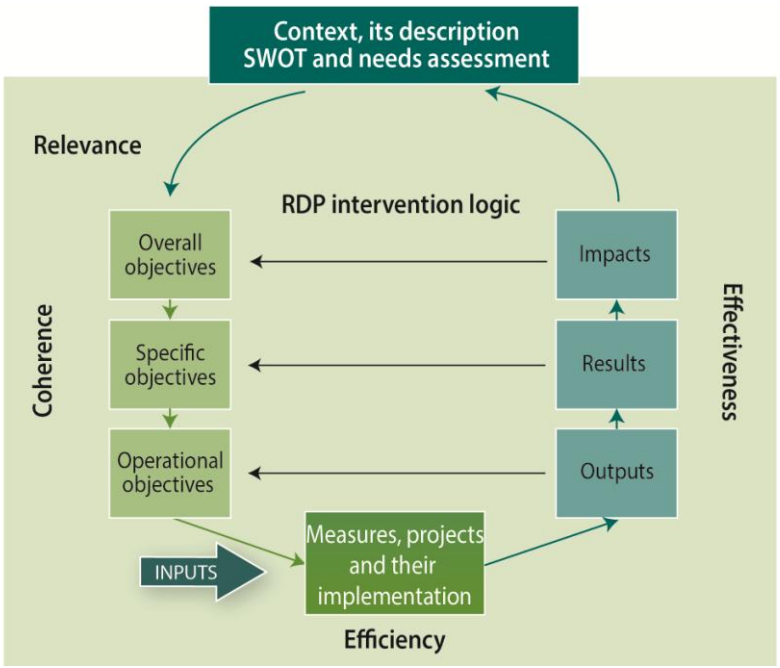
What is the intervention logic?

Appraising the intervention logic of a Rural Development Programme or any measure is an essential cornerstone[44] for its assessment. As outlined in the glossary of the Common monitoring and evaluation framework for rural development (CMEF) of 2007-2013, the intervention logic;

> "…*represents a methodological instrument which establishes the logical link between programme objectives and the envisaged operational actions. It shows the conceptual link from an intervention's input to its output and, subsequently, to its results and impacts.*"

The quality of the intervention logic, essentially its coherence, robustness and relevance, will be a major influence on whether the programme objectives are achieved. The CMEF, Guidance note B illustrates the positioning of intervention logic with Figure 3 below. Essentially the intervention logic frames the rationale for and means of intervention in relation to established programme needs.

Figure 3 Intervention logic of RDP and its context



Source: Helpdesk of the European Evaluation Network for Rural Development based on the CMEF 2007- 2013

The intervention logic is the framework of thinking and should be designed in a theoretically robust and yet practical way addressing the questions: (i) What are the needs being tackled and what are the interventions proposed to address them? (ii) Is there a justification for the proposals?

A well designed intervention logic must reflect the most important needs, framed by policy objectives and identified through the description of the situation in the programme area, SWOT analysis and needs assessment, as those to be addressed by the rural policy intervention within the time of the

[44]CMEF for rural development 2007-2013, Guidance note B

programming period in a timely and cost-effective manner and efficient use of the resources available. A hierarchy of objectives, ideally formulated in a SMART[45] approach, represents the policy response to identified needs at various levels:

- Overall objectives express envisioned long-term changes that stakeholders want to achieve in the programme area in relation to identified needs. These changes will ideally happen if the programme produces the expected impacts. Overall objectives are further broken down to specific objectives.
- Specific objectives are about necessary short-term changes that need to happen within the group of programme beneficiaries of various rural development sectors or geographical areas in order to achieve overall objectives. These changes will happen if the programme or its parts (axes, groups of measures and/or individual measures) generate the expected results.
- Operational objectives belong to the lowest hierarchical level and express the immediate changes which should happen at the level of programme beneficiaries to accomplish specific objectives. These changes will only happen if the expected outputs are created through the implementation of measures and operations selected and combined to enable the achievements of objectives.

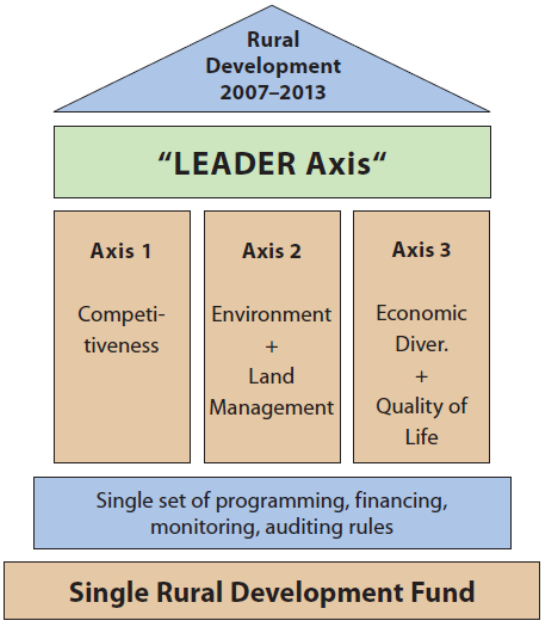
Because of the length of the Rural Development Programme period, it may be necessary to alter the intervention logic part way through the delivery, partly to respond to unanticipated changes and partly to learn from what is or is not working well in terms of planned effects.

What is the intervention logic of EU rural development policy

Each programming period of the European Union is guided by a common EU policy framework, which provides a high-level intervention logic for various national and regional policies. In the EU programming period of 2007 – 2013, rural development policy was guided by the Community strategic guidelines for rural development 2007- 2013[46]. The Community strategic guidelines define six overall policy objectives[47], which are transformed into overall objectives defined in a legal act[48] and implemented by means of four axes[49] and related measures. The architecture of the EU policy framework is shown in Figure 4:

[45] SMART : specific, measurable , achievable, relevant and time-bound
[46] Council decision (EC) 144/2006 on community strategic guidelines for rural development 2007 - 2013
[47] See Introduction: Scope of *ex post* evaluation
[48] Council Regulation (EC) no 1698/2005, Art. 4.1
[49] Council Regulation (EC) no 1698/2005, Art. 4.2

Figure 4 The architecture of rural development policy 2007-2013



Source: European Commission Fact Sheet2006

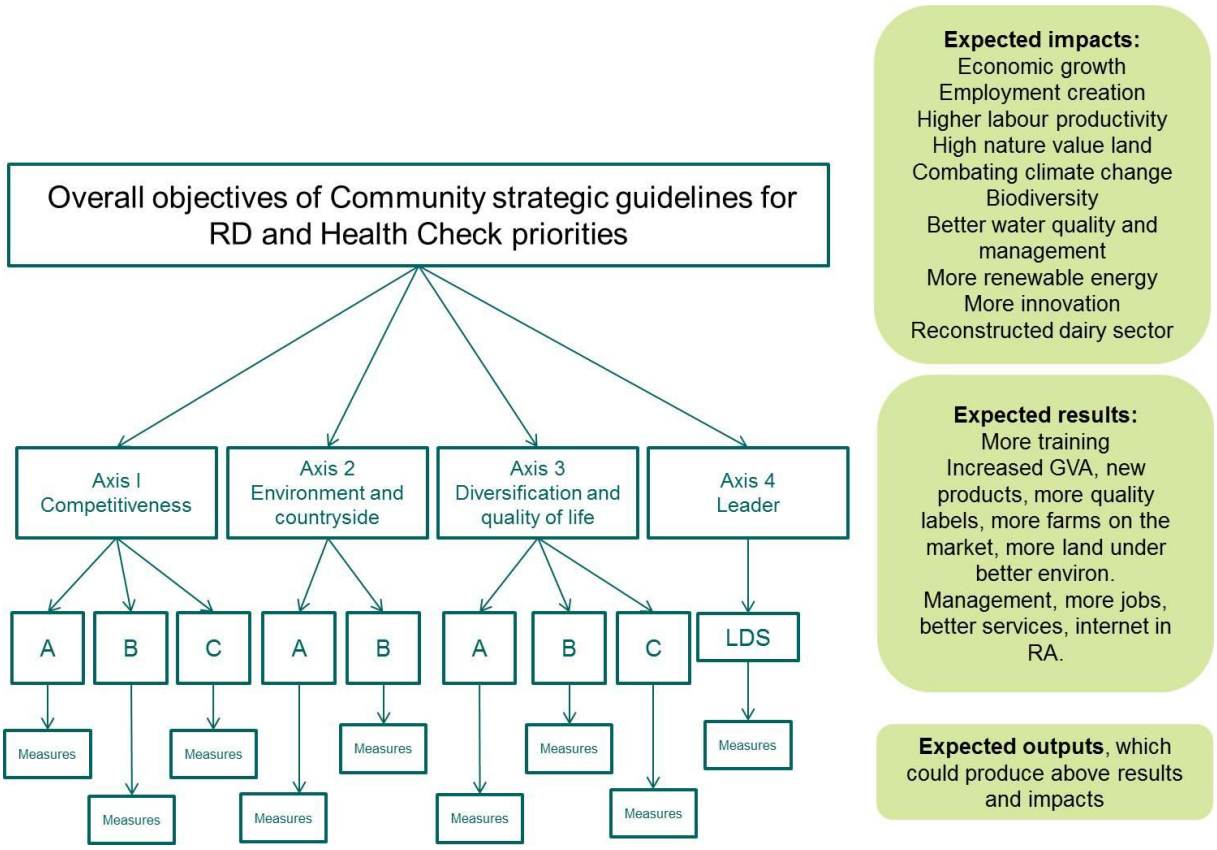
The overarching EU policy logic for the 2007-2013 RDP can be seen as recognising the centrality of rural areas multi-functionality while, at the same time, acknowledging that rural development as a form of policy intervention needs to go beyond the farm sector and embrace a wider range of rural actors. There was an explicit recognition that rural development, while acknowledging and respecting the need to develop the farm and forest sectors and ensuring that these sectors delivered important environmental public goods and avoided environmental bads (such as diffuse pollution), comprised a more all-embracing interest in sustainable rural development and the wellbeing and advancement of rural communities. The result of such a vision is a very long "menu" of possible measures, with Member States having a great deal of discretion as to which measures they adopted.

The original set of policy objectives was supplemented by the additional rural development priorities in a consequence of the Health Check of the Common agricultural policy reform in 2008, reflecting recent challenges for EU agriculture and rural areas: climate change, renewable energies, water management, biodiversity, innovation linked to all those above plus measures accompanying the restructuring of dairy sector[50].

EU rural policy objectives and Health Check policy priorities, as well as related expected outputs, results and impacts comprise the EU high policy level framework, which constitutes the basis for the design of programme intervention logic. At each stage of the evaluation process the intervention logic needs to be scrutinised critically, with the *ex ante* analysis asking whether the intervention logic is plausibly framed; the mid-term evaluation seeking whether the intervention logic appears to be robust and the *ex post* evaluation assessing whether the intervention logic was robust and effective with the benefit of hindsight. The EU policy framework is illustrated in Figure 5. More detailed links between policy objectives at each hierarchical level of the EU intervention logics for each axis and sub-axis are illustrated in **Part III:** Item 1.

[50]Council Regulation (EC) No 74/2009, The Health Check of the CAP reform

Figure 5 EU policy framework for rural development 2007 – 2013



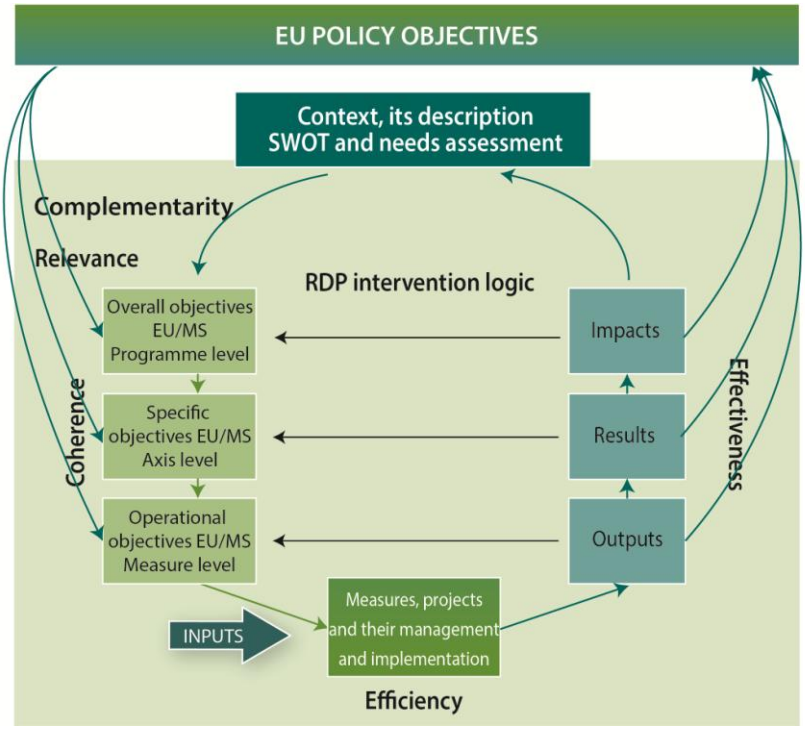
Source: Helpdesk of the European Evaluation Network for Rural Development

Although the regulation obliges Member States to respond to the overarching policy goals, at the same time, it encourages them to address the national specificities regarding rural development. Therefore at national and regional level, the EU policy objectives set up by the Community strategic guidelines and Health Check of the CAP are transferred into Rural Development Programmes, taking into consideration specificities of the respective Member States or regions. In this respect, the EU policy objectives and priorities encourage programme authorities to think how best to enhance competitiveness through restructuring and investment support, how to address environmental quality issues associated with productive land use and how to go beyond narrowly circumscribed notions of growth and address quality of life.

The review of the intervention logic must be set within both the Commission's structuring of the rural development challenge and the existing suite of support policies operated by the Member State. In all cases, a side-ways examination of other policies and measures seeking to achieve the same types of outcomes should ensure that the intervention logic for the RDP is complementary with other programmes and policies and does not duplicate existing measures.

The RDP intervention logic respecting the common EU policies as well as reflecting the programme areas' specific needs can be illustrated with Figure 6.

Figure 6 RDP specific intervention logic with common and programme specific elements



Source: Helpdesk of the European Evaluation Network for Rural Development

The intervention logic should be explicitly stated in the National Strategy Plan[51]. This National Strategy Plan brings together the social, economic and environmental situation analysis and the potential for development, as well as linking European rural development policy with national and regional programmes. If the intervention logic is not convincingly asserted in the National Strategy Plan, this should be flagged as a significant weakness.

The evolution of intervention logic over time

The Health Check agreed in late 2008 allowed "*a better response to the new challenges and opportunities faced by European agriculture, including climate change, the need for better water management, the protection of biodiversity, and the production of green energy. Member States will also be able to assist dairy farmers in sensitive regions adjust to the new market situation.*"[52] This created a need for modified intervention logic and/or smarter use of existing measures to attain these modified objectives.

Also events may happen that derail expectations. The European banking crisis in the last programming period reputedly had a large impact on the ability of rural businesses to draw down credit from commercial sources under the Leader axis. Global farm product prices have been very volatile. In those countries most seriously affected by the credit crisis, spending the allocated funds was really challenging. It does not mean that with respect to Leader, the local development strategy was ill-conceived, but simply that events had occurred which derailed expectations and unanticipated events such as the banking crisis (termed by the author Nassim Nicholas Taleb[53] as "black swans") could not reasonably have been anticipated. The result was almost certainly a lower degree of leverage because those with savings were almost the only people in badly affected countries who could engage in projects. Those in greatest need had least access to funds because the co-financing obligation could not be met.

[51] Council Regulation (EC) No 1698/2005, Art. 11
[52] Health Check of the Common Agricultural Policy
[53] Nassim Nicholas Taleb, 2007 The Black Swan: The Impact of the Highly Improbable, Penguin Books

1.2. What are the major challenges in the intervention logic review?

The biggest challenge with respect to the assessment of intervention logic is to go beyond the "visible" intervention logics set up in accordance with legal[54] and administrative requirements, using a theoretically robust lens to examine the logical coherence, theoretic rigour and practicability of the programme, axes and measures.

Various clusters of intervention logic

In this respect three main overarching types or clusters of intervention logic can be identified in relation to measures. Of course, a measure may be informed and shaped by more than one "logic" and therefore it is highly desirable to frame and subject it to critical scrutiny. These bundles of logic broadly correspond to Axis 1 (economy), Axis 2 (public goods/bads and regulatory compliance) and Axes 3 and 4 (quality of life and social objectives).

Economy related cluster of intervention logics

The first cluster of intervention logics relates to rural *economy*. The critical question is whether the proposed intervention(s) will contribute to sustainable economic growth. The specific economic rationale for the intervention is likely to vary greatly from measure to measure.

Some measures are designed to support business development and competitiveness. These are of several types:

- They can involve the development of human capital through training, early retirement schemes, new entrant schemes, etc.;
- They can support business restructuring by supporting the provision of fixed capital or new technologies, with capital support for investments which enhance business efficiency;
- They can support new infrastructure needed to support the adaptation of agriculture and forestry;
- They can support the development of food supply chains; and
- They can support of product quality improvements (including PDO, PGI and TSG designations)[55].

Following the identification of a policy need, some countries adopted the measure which supported the restructuring of semi-subsistence agriculture, including diversification, which can be seen as supporting the domestic economy.

In using the above measure, two rather different and separate intervention logics may apply:

- First, the more effective production of food for domestic livelihoods may target the subsistence needs of a relatively poor section of the population.
- Second, the subsistence farming practices may offer scope for up-scaling into commercial enterprise and may afford opportunities for the development of niche market speciality foods which might then perhaps seek designated speciality food status. Associated with this might be a need to develop new business skills.

Another type of economic rationale can also be used to justify Axis 2 measures. Economists recognise the widespread problem of externalities in agriculture and the tendency for overprovision of

[54]Rural development priorities as laid down in Community strategic guidelines for rural development 2006/144/EC and RD objectives, axes and measures as laid down in Council Regulation No 1698/2005

[55]PDO: Protected designation of origin
PGI: Protected geographical indication
TSG: Traditional speciality guaranteed

environmental "public bads" like water pollution and the under provision of public goods such as biodiversity. Externalities are negative or positive spill-over effects arising from farming activity. Axis 2 explicitly offers measures to enable farmers to be compensated for the provision of public goods and to be persuaded to undertake measures to reduce the impact of negative externalities such as water pollution. We should be seeking an intervention logic that recognises and where possible values these goods or bads and shifts production towards a more socially optimal outcome.

Another type of economic rationale is the idea to strengthen rural economies. The Axis 3 measures relate principally to diversification of rural economies and support for micro-enterprises, including tourism. Here the implicit idea is that diversification helps resilience and may support faster growth than in situations where there is overdependence on mainstream farming. Arguably, the support of upgrading heritage and village renewal can also be seen as enhancing the fixed infrastructure in which diversification can take place. The rationale here is identical to that for the farm sector, except it extends more widely to providing infrastructure for the whole rural economy.

Public goods/bads and regulatory compliance related cluster of intervention logics

A second cluster of intervention logics relates to regulatory compliance. The regulations in question may be national or EU-wide. There are two main areas where regulatory compliance is supported in the RDP and a third where regulatory compliance can be seen as a possible motive. Farmer and forest owners' compliance with EU Directives to support habitats or particular species conservation (the Natura 2000 Directive and the Birds Directive) or water quality (Water Framework Directive) have been supported under the Axis 2 measures. In addition, the selection of species and habitats to be supported can be framed by their presence in other regulatory measures such as Biodiversity Action Plans. Support might also be given for non-binding regulations such as those involving sustainable forest management or compliance with other EU standards such as organic farming. In these cases, the intervention logic is partly to support regulatory compliance and partly to support public goods or avoid public bads. But, in addition to supporting regulatory compliance with important environmental policies at European level, some support is also given to ensuring compliance with food quality schemes. Such measures may be explicitly related to quality regimes or labels.

Quality of life and social objectives related cluster of intervention logics

A third cluster of intervention logics relates to predominantly social objectives. The provision of basic services in rural areas and the conservation and upgrading of rural heritage can be seen not only as investments to support economic development, but also as a means to support a better quality of life, which cannot, according to many experts, be reduced to a narrow economic metric. Indeed, legal acts[56] state that rural development policy should support the reduction of social and spatial inequalities and address the challenges faced by disadvantaged and depopulating rural regions. Such investments may be associated with a desire to make more attractive living space to encourage younger people to stay in an area. Training for leadership and support for capacity building may also be either economic or social in emphasis.

A deeper exploration of the intervention logic may reveal an interconnection between economic and social logics. These may expose what may be contestable intervention logic. How are the weightings between different policies assessed in the absence of a common metric? In a sense, this reveals that multiple strategies are possible: for example, one programme may place emphasis on economic development as a means to enhance wage rates and employment prospects and thereby to reduce inequalities, whereas another might seek social interventions that explicitly reduce the disadvantage faced by poor people. These perspectives may be rooted in different ideologies, which it is better to be explicit about than to "sweep under the carpet".

[56]Council Regulation (EC) No 1698/2005, recital 46

Among very important intervention logics which belong to this cluster are those linked to Leader approach. Leader is not only the axis of rural policy that implements rural development measures through local development strategies - LDS (mini-programmes, which have their own intervention logic). Leader also represents a method through which specific EU rural policy interventions are implemented, at the same time generating and enhancing social capital and hopefully triggering changes in:

- individual capacities, namely concerning knowledge, attitudes and behaviours, as well as
- organisational and institutional capacities (strategies, rules, routines and culture), as well as
- societal capacities, namely inter-organisational cooperation links and network relationships as well as normative capacities (cultural and regulatory framework determining the development of new options for regional governance in rural areas).[57]

The specific substance of Leader – the LDS and generation or enhancement of social capital causes complexity in evaluation, not least because Leader is about both: processes and impacts/outcomes. There are several aspects, which need to be taken in consideration when looking at the RDP intervention logic with lens of Leader:

- Through local development strategies, Leader contributes to RDP, complementing and reinforcing the actions undertaken in Axes 1, 2 and 3, and thus contributing to RDP outputs, results and impacts. This contribution is reinforced with the assumption that the endogenous, participatory, partnership-based and integrative approach involves more beneficiaries and produces more sustainable results than the mainstream elements of programme delivery. It is said that the main Leader added value, the strengthening of the social capital of the area, may produce desired outcomes in less direct but ultimately more sustainable ways[58].
- The explicit emphasis on strengthening the local communities' governance potential has been introduced in the intervention logic of Leader for the first time in the period 2007-2013. The Commission Guide for the application of the Leader Axis[59] associates the term governance with the capacity of Leader to bring together a broad range of public, private and civil society partners, to forge a sense of identity, and to increase the local management and project development capacity of rural communities.
- The encouragement of innovation hinges on the latitude and flexibility that Leader is allowed to have within the national or regional framework. Innovation should be understood in a broad sense, and be simply seen as a way of finding new solutions to an area's needs.

The above induces two types of impacts which should be taken in consideration in RDP intervention logic assessment in relation to Leader:

- Impacts of Leader on the whole RDP implementation (all four Axes) are measured against the core objectives of rural development policies.
- Additional impacts of Leader that can be expected on the set-up (rules, structures and processes) of multi-level regional governance, local governance and ultimately on rural policy concepts; this impact would be attributable mainly to the Leader method and its added value, determined by:

[57]The theory of change of the Leader approach is extensively described on pages 56-60 in: Metis et.al. (2010): European *ex post* evaluation of Leader+, Vienna.
[58]European Commission, DG Agri Unit G4 (2004): Methods for and Success of Mainstreaming Leader Innovations and Approach into Rural Development Programmes. Final Report.ÖIR Managementdienste, Vienna; p.57.
[59]DG AGRI Guide for the application of the Leader Axis of the Rural Development Programmes 2007-2013 funded by the EAFRD

- outputs achieved through rolling out the local development strategy and funding Axis 4 projects, plus
- additional results and impacts achieved by investments in human and social capital, plus
- changes in multi-level and local governance for rural development, minus
- results and impacts which would presumably have been achieved by implementing the same type, number and volume of projects attributable to measures under Axis 1, 2 and 3 of the RDP.

The three different clusters of intervention logics may look rather distinctive, but in practice they are likely to be interwoven, especially through the Leader process. For example, regulation may support consistency of quality of a product and lead to positive economic outcomes. Enhancing quality of life through capacity building may support community vibrancy and also drive economic outcomes. Collaborative natural resource management to deliver enhanced compliance may build social capital and enhance quality of life. These win-win or even win-win-win outcomes are not necessarily serendipitous. They can be framed by an intervention logic which seeks to rationalise the measure in such terms, by having multifaceted intervention logic. Notwithstanding the transaction costs and monitoring difficulties associated with more complex intervention logics, we should see such win-win outcomes as desirable and seek to establish whether measure choice matches the multiple intervention logics.

Without explicitly framing the intervention logic in each case, CMEF Guidance Note E indicates the rationale for introducing the measure. The interconnectivity to which we refer above can be considered by reference to the different clusters of intervention logics and can usefully be separated into primary intervention logic and any subsidiary intervention logics. Examples of intervention logic for various groups of measures are in Table 3 below.

Table 3 RDP Measures and associated intervention logics

Measure	Primary intervention logic	Subsidiary intervention logic(s)
Measures 111-115: Training support and new entrants.	Economic performance	Non-market goods and bads/QOL
Measures 121-126 Modernisation of holdings, etc.	Economic performance	Non-market goods and bads/QOL
Measures 131-133, 142 Standards quality schemes and producer group promotions.	Economic performance	QOL
Measure 141 Semi-subsistence holdings.	Economic performance	QOL
Measures 211-212 Natural handicaps payments.	Non-market goods and bads	QOL
Measures 213, 224 Compliance with Natura and WFD.	Regulatory compliance	
Measure 214 Agro-environmental payments.	Non-market goods and bads	Economic performance
Measure 215 Animal health and welfare.	Economic performance?	Non-market goods and bads?
Measures 216 and 227 Non-productive investments.	Non-market goods and bads	QOL
Measures 221-226 (excluding 224) Forestry support.	Economic performance	Non-market goods and bads

Measure	Primary intervention logic	Subsidiary intervention logic(s)
Measures 311-313 Diversification	Economic development	Resilience/QOL
Measures 321-323 Services and village infrastructure	Quality of Life	Resilience/Economic performance
Measures 331,341, 41, 421, 431 Quality of life and Leader	Quality of life	Resilience/Economic performance

As an example of intervention logic at programme level we can examine the Scottish Rural Development Programme 2007-2013. It asserts: Two key principles underpin the priorities for the 2007-2013 SRDP:

- The first is sustainability whereby measures achieve complementary outcomes – whether social, economic or environmental – and avoid net damage to the cultural and historic environment.
- The second is the need to correct for market failure and deliver outcomes that are for the benefit of rural communities and the wider population.

The Scottish Government argues that their choice of measures is guided by clear principles:

- There should be evidence to justify funding.
- Support should be targeted at areas, sectors or communities where there is a demonstrable and significant need.
- There should be no duplication through other funding streams such as the EU Structural Funds; indeed measures should complement other initiatives and be consistent with local authority plans and local economic strategies.
- Funding should be the most appropriate solution, either in terms of building capacity or supporting projects.
- Funding must deliver outcomes that would not happen otherwise and make a significant and, wherever possible, measurable improvement.'

The intervention logic should be assessed for each measure against these demanding principles.

In contrast to Scotland's twin emphasis on addressing sustainability and market failure, the Finnish Rural Development Programme articulates the same need to address sustainability but rather differently emphasises the need to develop and support new rural enterprise and economic diversification and strengthen local initiatives to improve the quality of rural life. These differences between countries and their different underlying intervention logics should be teased out by the evaluator who should be addressing the specific programme emphasis in the review of the intervention logic.

1.3. What should be reviewed in relation to intervention logic in the *ex post* evaluation?

The review of the intervention logic is an integral part of the *ex post* evaluation of a Rural Development Programme. Given that there is Member State discretion in the weighting between axes, this needs to be justified in the appraisal of the intervention logic, albeit within limits set by the Commission guidelines.

The assessment of intervention logic needs to go beyond the administrative requirements and pure theoretical appraisal using a robust lens to examine the practicability of the programme, axes and measures. The assessment of intervention logic shall be reviewed for its robustness (relevance and coherence), unintended effects and efficiency. This appraisal will allow the uncovering of any structural weaknesses of an intervention logic that lie behind the true policy needs, to better understand the evaluation results, contribute to the policy learning and improvement of the design of future programmes.

Assessment of the relevance of intervention logic

The assessment of relevance looks at the intervention logic´s ability to respond to the needs and their evolution, responding to the changing environment as well as changing policies (e.g. the EU Health Check of the CAP) over the life cycle of the programme. The review of intervention logic´s relevance should:

- Assess whether pre-defined programme objectives were relevant and whether activities designed to meet these objectives were most suitable (if not what would have been more appropriate).
- Appraise whether there was a logical, theoretically well-grounded explanation of the rationale for policy intervention,
- Help to understand if the selection and composition of measures was designed for a wide group of potential beneficiaries or more narrowly framed to benefit particular groups and, if so, why,
- Assess whether Leader was designed in addressing specific needs of rural areas in relation to follow the seven Leader approach principles,
- Appraise the extent to which the intervention logic was able to evolve along changing needs and policies (e.g. Health Check) of the programme area and address them properly,
- Look at extent to which the intervention logic was able to address the identified needs and changes occurring (e.g. Health Check) across the programme implementation,
- Draw important lessons in relation to practical applicability of individual measures in addressing the needs.

Assessment of the coherence of intervention logic

The assessment of the intervention logic´s coherence looks at the robustness of linkages between objectives and the outputs, results and impacts produced with the means of inputs: measures and allocated funds. The intervention logic should be a “cascade of coherence” where the overarching objectives are achieved by the four Axes and their own internal logic, which is then mirrored by the logic of the suite of measures.

The assessment of coherence is looking at vertical coherence within the hierarchy of objectives broken down to measures and allocated funds and from there, along the chain of outputs, results and impacts produced with these funds. The assessment of intervention logic also looks at horizontal coherence between objectives at the same hierarchical level as well as in relation to level related

generated effects (outputs, results and impacts). For the programme evaluation, it is very important to look at horizontal coherence between objectives which tackle various types of interventions: competitiveness, environment and quality of life, and see the conflicts, gaps or synergies between them[60]. The assessment of the intervention logic´s coherence will:

- Provide, with hindsight, evidence of whether the thinking behind the rationale for policy intervention was coherent and robust,
- Provide an appraisal of the extent to which higher level objectives could be reached via achievements of lower level objectives,
- Assess whether the inputs of selected measures and their composition generated outputs which led to produced results and impacts,
- Assess whether programme objectives were achieved due to outputs, results and impacts produced by the selection and combination of measures or allocated funds,
- Appraise to what extent were the objectives at the same level mutually reinforcing,
- Look at how the selection and combination of measures and funds have contributed to the antagonism or synergy between various objectives at the same hierarchical level,
- Assess whether the response of beneficiaries to measures (uptake) was sufficient enough to achieve objectives,
- Conduct the appraisal, as to whether the Axis 4 outputs can produce results and impacts in relation to the added value of Leader,
- Look whether the objectives linked to governance, building capacities and innovation in relation to the Leader approach were able to be achieved with the means of inputs, selection and combination of measures and outputs received,
- Appraise whether the programme´s intervention logic was able to contribute to the achievement of the EU rural policy objectives.

Assessment of the unintended effects of intervention logic

As well as delivering desired outcomes and impacts, every intervention also creates the distinct possibility of unintended effects on society, economy and environment, which could either be positive and negative. The *ex post* review of intervention logic must assess whether the selection and composition of measures might have created foreseeable unintended effects in the programme area, both positive and negative to the intervention and take these unintended effects in consideration when conducting the evaluation.

For example, overlooking any significant unintended negative effects in the intervention logic design can result in the “healing” of one problem in one sector generating an “unwanted illness” in another sector (farm diversification measures supporting new tourism could either displace an existing local tourism business or be additional or even complementary to it).

Assessment of the impact of intervention logic on programme efficiency

Efficiency is the relationship between cost and achievement. An intervention logic that is well designed and properly thought through should facilitate or lead to high level of programme effectiveness and efficiency.

[60] Guidelines for the *ex ante* evaluation of 2014-2020 RDPs, Chapter 2.3 Intervention logic, Evaluation Helpdesk, 2012

The assessment of the intervention logic should appraise:

- Whether the allocated inputs were able to produce high levels of outputs,
- Whether the generated outputs could have been produced with less inputs,
- Whether another measure (or set of implementation practices) would achieve the same outcome more cost-effectively.

In summary, an effective intervention logic is likely to result in efficient use of resources in the programme.

Experiences with intervention logic and its assessment

Exploring the intervention logic should not be a "box ticking" exercise involving a description of how the selection of the measure was conceived and why it was chosen. It can and should go much deeper than that.

In each case the evaluator must be alert to shaky and ill-founded logic. Such an intervention logic can arise from the power of sectional interests to steer benefits towards themselves (sometimes termed confusingly 'rent-seeking'), the desire to appease particular interest groups, the failure to anticipate land manager responses, the confusion of policy logics (for example by a muddle of logics including economic and regulatory compliance motives). In each case, we should be seeking to find an intervention logic which avoids preferential treatment of non-deserving groups of beneficiaries.

There is a big step from the highly plausible intervention logic articulated in the Commission's reasoning about rural development and the practices in some Member States which reflect the power of strong interest groups to draw down benefit in what Schneider and Ingram (1997) see as regressive policy designs. According to the authors regenerative policies are those where the policy means delivers to a publicly constructed agenda; and it is such an approach that we should be seeking. Degenerative policies are those captured by particular interests; essentially where sectional interests capture the policy mechanism to serve their own interests, rather than the public interest[61]. Given the breadth and scope of the Rural Development Programmes and the power of some sectional interests in Europe, regression to degenerative policy designs is an ever-present threat.

Another example of ill-founded logic is where additional resources are being allocated to a sector to try to achieve particular outcomes, but the underlying rationale for the intervention is not strong enough to justify such a use of resources. The most likely causes of such interventions are path dependencies (still doing something that should have been laid to rest a long time ago, often because it is administratively easy, or keeps key stakeholders happy) or rent-seeking behaviour (drawing down public money) by powerful lobbies in the policy process.

1.4. How should the intervention logic be reviewed or assessed during the *ex post* evaluation?

One recommends a two-stage process to assess the intervention logic. This will enable an exploration of the intervention logic of the intentions and of any Health Check generated adjustments; and second, in the light of evidence found in the evaluation to revisit the intervention logic and assess its robustness.

The first stage should take place early in the *ex post* evaluation as the initial screening of the actual programme intervention logic in the light of pre-identified types of intervention logic. At this stage, we should be looking at what is meant to be achieved with the intervention logic and how it is proposed to be achieved (e.g. selection and composition of measures or operations around various Axes). The

[61]Schneider, A.L. and Ingram, H., (1997) Policy design for democracy, University Press of Kansas, Lawrence Ks.

assessment also includes looking at the intervention logic´s evolution along with the implementation of the programme, recognising the need to factor in Health Check modifications.

The second stage revisits the intervention logic towards the end of the *ex post* evaluation exercise, after the evaluation methods has been applied and enables the assessment of various measures and the entire programme. At this stage, we should look if the proposed composition of intervention logic was able to achieved what was meant to be achieved and, if not, why.

Key steps in the assessment of intervention logic

The first task is to ensure that the evaluator has a clear grasp of the intervention logic (See **Part II**: sub-chapter 1.3). Second, we suggest the means by which the intervention logic can be explored. The evaluation of the intervention logic should be at programme level, axis level and measure level. There are often several different intervention logics behind an individual measure. Where there are several logics, it is incumbent on the evaluator to explore them all. Below, we suggest the means by which to do this.

Although it should be highlighted that the procedure applied in the intervention logic appraisal depends on the evaluator's decision, it is recommended to undertake a twin-track approach. This approach is conducted in two steps:

- First, the programme as a whole and its component parts should be subject to desk-based scrutiny where the evaluator uses theoretical and practical knowledge to assess the relevance, the internal coherence of the programme, its axes and the selected measures, the ability to achieve objectives, the intended and unintended effects and efficiency.
- Second, the evaluator can use two facilitated workshops to check out the evaluation team's judgements against those of a group of key informants. We first describe the stage-by-stage approach and then explore how the workshops can offer a form of triangulation to check out any interim judgements arrived at by the evaluation team.

Desk-based scrutiny of intervention logic

In the desk-based scrutiny, the evaluator should be familiar with the scientific and grey literature on rural development policy analysis. This should be screened for examples of policy successes and failures, paying particular regard to the potential issues of time and space-specificity in policy success or failure. There are some obvious lessons to be derived from this extensive literature: for example, farmers strongly focused on production are less likely to adopt environmental measures than hobby farmers; dairy farmers are less likely to diversify their businesses than many other types of farmer. What really matters is that the programme as a whole and the suite of measures delivers appropriate impacts. And behind a good or poor impact can lie the difference between a well-thought through intervention logic and a less well thought through logic.

The choice of measure is only part of the picture. The intervention logic appraisal must ask whether the uptake is by those whose response would be such as to maximise the net public benefit, whether they are likely to be responsive to the measure and what might impede such a process. In this respect, there is a need for scrutiny of the relationship between the implementation of a measure and the results of a measure on the intended target. Has it been taken up as expected? Has it delivered the intended impacts? The chapter 4 on methods identifies the suite of models and approaches that can be drawn on.

It is all too easy to take a rather blinkered and analytically closed view of the part of the land use system targeted, but in practice as an open system it is impacted by many possible changes. For example, succession policy and uptake of new entrants' measures may be affected by fiscal policy

relating to inheritance tax that has nothing to do with farm policy. Time and again, external influences alter anticipated uptake.

As programmes have built on previous programmes, so there ought to be a sign of an institutional memory in the appraisal of the situation and in the intervention logic. Has there been an experience of good uptake of measures in previous programmes? Were there differences, say on, different types of farmland, or amongst different groups of farmers (tenants versus owner occupiers; dairy versus beef, large versus small, etc.)? Is there scientific or grey literature exploring this issue? Did recognition of these questions inform the choice of measures and frame the intervention logic? Quite clearly, it should.

In practice, the evaluator should be addressing the intervention logic top-down from programme to measure. However, often an individual evaluator comes with a particular view of the world or interpretation of the legitimacy of the intervention. It is an obligation on the evaluation team to review the intervention logic in a dispassionate but theoretically informed way. This should ideally be a multi-person task if conducted in-house by the team.

The six intervention logics below should frame the desk-based appraisal of intervention logic at programme, group of measures or measure level. This is essentially an in-house activity to be conducted by the team of evaluators. It should draw on their diverse knowledge and experience and be framed in an understanding of the theoretical rationales for intervention. These logics also relate to the three clusters identified above in chapter 1.2:

Intervention logic 1 - Business development, including succession (Axis 1)

Business development, including succession, is addressed by a raft of measures under Axis 1 relating to human capital development including vocational training, early retirement, new entrants, use and development of advisory services and a second group of measures relating to physical capital developments, restructuring, including infrastructure development, adding value to products, recovery from natural disasters including prevention, the promotion of innovation and co-operation in the development of new products processes and technologies.

One expects the evaluator here to scrutinise the intervention logic in the light of the analysis of the Member State's socio-economic situation and explore the choice of the selected measure and the policy means by which it has been implemented. This scrutiny includes the analysis of intervention logic against the overarching intervention principles of sustainability, economy and social development, followed by a measure-by-measure appraisal of intervention logic against public and research findings on the success of similar measures. For example, the issue of new entrants has been widely explored at European and Member State level, as have retirement schemes. The aggregate evidence in academic and policy evaluations is ambivalent about their success, and suggests that their success has varied with scheme design. Where there is substantial evidence that policy interventions of a particular type have not always generated positive results, the intervention logic particularly merits careful scrutiny. Was the measure a product of *de facto* "rent-seeking" by vested interests or does it address a real policy problem?

Intervention logic 2 - Business efficiency and competitiveness (Axis 1)

Business efficiency can be enhanced by enterprise level improvements which improve technical performance and business efficiency or by collective endeavours either by groups of farmers operating collaboratively or by different actors collaborating along supply chains. A critical question to explore in appraising the intervention logic of such support measures is why such measures are not happening without intervention. Some types of investment might be expected to contribute to beneficial environmental outcomes for which the land manager is not rewarded. Where public goods are enhanced or public bads reduced the intervention logic is clear. Where the objective is collaboration

there are a number of possible intervention logics from mutual learning to enhanced technical and economic efficiency. Here the intervention logic is based around a combination of enhanced competitiveness through human resource development and infrastructural and technical investment support.

Intervention logic 3 - Environmental quality enhancement (Axis 2 and maybe 3 and 4)

There are at least two intervention logics with respect to environmental quality. The first relates to regulatory compliance. The second relates to economic efficiency.

The primary requirements for regulatory compliance arise from the implementation of measures to deliver compliance with the measures of the EU Water Framework Directive (WFD) and the Natura 2000 and Birds Directives on habitats and species. The principal driver of compliance with the WFD is the recognition that water quality is often compromised by diffuse pollution arising from agricultural sources. However, it is incumbent on those designing policy to have a clear idea of where diffuse pollution is coming from. Could it be septic tank leakage from non-farm properties? Are we dealing with nitrate or phosphate pollution or both? Is the problem one of emissions from buildings and yards or from fields? Although the WFD articulates a disproportionality principle (that is an intervention is only required if the cost is not disproportionate to those applying it), is this principle effectively applied?

The compliance with Habitats and Species Directives hinges around the fact that farming systems, in particular those which retain semi-natural habitats, deliver these environmental public goods as co-products of farming practices. Where there are market or business pressures to change farming systems in ways which compromise the species or habitats, there is a need to lever compliance. This is the basis for a proportion of the RDP spending in many countries.

It should not be forgotten that there is also strong intervention logic to reward providers of public goods under the Provider Paid Principle and to invoke the Polluter Pays Principle. The economic logic argues that an unrewarded externality is underprovided and the producer of an un-taxed negative externality tends to produce beyond the social optimum.

Intervention logic 4 - Diversification and resilience

The desire to support greater resilience is incontestable. The challenge is how to create greater resilience. This resilience can be measured by the ability of an entity (it could be a business a co-operative venture or a community) to withstand adverse shocks and recover. The challenge for the evaluator is how to effectively scrutinise the intervention logic around the enhancement of resilience.

One important issue is in relation to diversification and resilience is potential displacement. Some RDPs have tended to support farm diversification into alternative enterprises such as equestrian enterprises and tourism, whilst mainstream (non-farm) equestrian and tourism enterprises are not supported. Unless the new business attracts different customers, a degree of displacement is likely. The chat board of the National Rural Network in England clearly shows that some equestrian enterprises that were not eligible for RDP funding felt that they were losing trade to newly supported RDP beneficiaries. This implies a risk of displacement rather than additionality and may actually reduce rather than promote aggregate resilience.

Intervention logic 5 - Quality of life

The intervention logic with regard to measures to support the quality of life seems to be rather uncontroversial: interventions which enhance the business environment and support livelihoods or support the wider qualities of a place - the 'liveability' of an area are justifiable on these grounds. However, it is much more questionable as to whether the measure supporting either livelihoods or liveability was selected on the basis of surveyed need or active support seeking by a small number of

people with rather specific interests, whose drawdown of support is essentially to support their rather narrow interest, rather than the wider public interest.

One of the particular challenges with intervention logics relating to quality of life is the intended list of beneficiaries. This should be scrutinised carefully. Is there evidence that particular groups of beneficiaries are likely to benefit or have benefited and does the level of benefit derived suggest that this was a good place to direct support?

Intervention logic 6 - Partnership-based approaches

The logic of the Leader approach and also for the capacity building measures that deal with group-based activities is often that partnerships deliver better outcomes and impacts than individual agency work. This is perfectly plausible and sound intervention logic. But the term partnership covers a multitude of organisational forms, some of which are very inclusive and others of which tend to be dominated by one particular group. They vary in their decision-making capacity. Partnerships also vary enormously in their effectiveness. Some pull key players together and share resources and undoubtedly achieve more. Others become time-consuming 'talking shops'. A representative group of partners knows how well they work; so too would a relatively small sample of those who should be benefitting from its actions. They need key players and not token representatives to work, but they also need to think through the intervention logic of how and where they direct their resources.

The use of facilitated workshops to triangulate findings

It is recommended that the evaluation team uses a two-stage facilitated workshop using expert informants to test the robustness of the intervention logic. The experts should be familiar with the programme, and should have expertise in four main areas: economic development/competitiveness in rural areas, public goods/bads associated with rural land use, especially biodiversity and water quality, regulatory compliance and quality of life in rural areas. It is crucial that a plurality of views is accommodated; and, where possible, experts who are not affiliated to sectional interests should be selected. We recommend a cascaded approach where first the overarching logic of the programme is confronted; then the SWOT is critiqued and finally the measures and implementation approaches are subjected to scrutiny.

It is suggested to hold the first workshop relatively early in the period in which the evaluation takes place, when the team has checked the intervention logic against measures and it is the overall shape of the programme and selection of measures and the scrutiny of their intervention logic that should be the focus. The second workshop should take place with key informants more as a wash-up meeting, where, after the evaluation methods have been applied and the full array of indicators is available, a facilitated workshop probes the key informants for explanations behind more successful and less successful measures.

Consideration of external factors that may be the cause of unexpected effects

External factors have considerable capacity to generate unanticipated outcomes in terms of policy uptake. For example, farmers are understandably slow in initiating measures to take out land from production where commodity prices are buoyant. A strong currency vis-à-vis the Euro can also make a big difference on farm profitability. So intervention logics cannot be deemed successful just because an impact indicator shows a distinctly positive change. For example, those with farm woodlands may be more inclined to rehabilitate them when firewood markets have expanded because of new non-farm policy measures such as the Renewable Heat Incentive introduced in the UK in 2013 to encourage adoption of renewables in space heating. Or oil prices may have gone up or down and influenced the demand for wood fuel. Thus, the *ex post* assessment of intervention logic must recognise the scope for unanticipated events to derail expectations. This does not make the intervention logic flawed, but it

sets outcomes in a context of the likelihood of needing to modify intervention logic in future programmes. That understanding should be stored in the institutional memory.

1.5. Specificities with respect to TA, NRN

Technical assistance

Member States can allocate up to 4 % of the total amount of each Rural Development Programme budget to finance technical assistance in the form of preparation, management, monitoring, evaluation, and information and control activities of programme assistance[62]. Technical assistance covers a wide variety of activities and although this activity is not formally formulated as intervention logic and covered by performance indicators, the evaluator should look at the logic of using the funds dedicated to technical assistance similarly to other RD interventions, since the same evaluation principles should apply for the technical assistance as for the programme.

In the assessment of the TA-related intervention logic the evaluator should see if clear intervention logic to the TA can be identified and, if clear objectives have been set up in relation to TA, what activities have been implemented and which funds have been allocated. The evaluator should also appraise whether the overall package of TA interventions was appropriate to deliver expected outputs and results in relation to overall challenges associated with programme preparation, monitoring, administrative support and evaluation and control[63]. Any evidence of a SWOT analysis of the administrative machinery to deliver the programme might be seen as positive evidence of recognition that the RDPs were complex and multifaceted in ways likely to create a need for TA.

As in the case of programme intervention logic the evaluator should look at the relevance, coherence, effectiveness of the TA, also using the examination of TA interventions in the *ex ante* and mid-term stage of the programme. In many cases, there were major challenges in the use of indicators in the assessment of the TA. Given that these issues were flagged at EU level and in many cases will be self-evident from a rudimentary examination of the MTEs, the extent to which TA was used to address problems and provide solutions to problems flagged in the MTE should be explicitly examined. Equally, the extent to which the Health Check created new TA needs and the extent to which those needs were met should be scrutinised.

The most appropriate means in the assessment of TA intervention logic is a combination of documentary analysis and key informant interviews (of the Managing Authority staff including Axis, measure and scheme managers) around a structured set of questions relating to their perceptions of TA needs and their judgement of the efficacy and utility of TA responses. Key documents relating to minutes of meetings addressing TA needs, requests for TA and records of contracts or other means used to deliver TA provide an evidence base. There is a strong case for interviewing those who made demands on TA and asking for the user's assessment of the utility of the TA in addressing problems.

National Rural Networks

Legal texts in the programming period 2007-2013 did not lay down common objectives for national rural networks (NRN); therefore most of them do not contain explicit intervention logic. Instead, the accompanying action plan is composed of a list of activities and a time-plan; hence, acting as operational objectives.

However, the European Commission has published the National Rural Network fiche in 2007. The fiche provides a series of useful recommendations in terms of structure of the network, allocation of human resources, using common networking tools and the action plan preparation and sets up minimal requirements for duties of National Rural Networks**,** stipulating among others the following:

[62]Council Regulation (EC) No 1698/2005,Art.66.1
[63]Council Regulation (EC) No 1698/2005,Art. 66.1

- facilitating an exchange of expertise at Member State level;
- supporting implementation and evaluation of the rural development policy;
- securing and coordinating the information flow between the local, national and European level.

The above requirements can be used as a lens when examining and justify the NRN's action plan and the intervention logic behind in the *ex post* evaluation.

In this respect, the task of the *ex post* evaluation will be to review the NRN intervention logic, if it exists or, if it does not exist, to establish the NRN intervention logic based on:

- the above common requirements and/or existing programme-specific objectives,
- the activities of the action plan,
- achieved outputs, results and impacts.

This established intervention logic should then be assessed for:

- its relevance in relation to the context of NRN interventions,
- its coherence looking at vertical and horizontal linkages between objectives and achieved outputs, results and impacts,
- its effectiveness, appraising whether the objectives or the above requirements of the NRN fiche have been achieved with the means of the action plan, and whether the NRN was able to contribute to RDP and EU overall objectives,
- effective use of the available funds in conducting activities of the action plan.

Wherever possible, the *ex post* evaluation of the NRN intervention logic should take in consideration existing studies undertaken with regard to the effectiveness of the NRNs.

1.6. "Dos and don´ts" in relation to assessment of intervention logic (based on experience)

DOs	DON'Ts
Assess the intervention logic as a three level process: of programme, axes and measures.	Consider the evaluation of the intervention logic as a box ticking administrative exercise.
Go beyond simple administrative compliance (is an intervention logic described?) to consider the possible intervention logics in relation to the theoretical rationale for intervention.	Focus only on the description of intervention logic when reviewing it.
Appraise the evolving or actual relevance of intervention logic with respect to needs and policies.	Look at the intervention logic without linking it to actual or evolving needs.
Check that the measures would qualify as regenerative rather than degenerative under the Ingram and Schneider classification of policy interventions (See **Part II:** Chapter 1.3).	Ignore the influence of particular political interests in the construction of intervention logic.
Assess the vertical and horizontal coherence between objectives and in relation to inputs, measures, outputs, result and impacts.	Assume that the stated intervention logic is always robust.

DOs	DON'Ts
Ensure that the primary and secondary intervention logics of each measure are thoroughly explained.	Focus only on direct links between measures, axis and overall programme objectives.
Assess the ability of intervention logic to achieve policy objectives with the means of selected measures and allocated funds.	Assume that objectives are achieved with any compositions of measures and funds.
Appraise the positive and negative unintended effects of the intervention logic.	Overlook unintended effects of the interventions.
Assess the efficiency of inputs in relation to generated outputs.	Overlook linkages between inputs and outputs in terms of cost efficiency.
Use a two-stage process in the assessment of intervention logic: at the beginning of *ex post* (before methods have been used) and at the end of *ex post*.	Conduct the assessment of the intervention logic only at the beginning of *ex post* evaluation.
Conduct desk research and use facilitated workshops with key informants twice during the *ex post* evaluation (at the beginning and at the end of *ex post*).	Conduct only the desk research when appraising the intervention logic.
Make sure that the distinctive features of the RDPs such as the NRNs, technical assistance are still considered fully with respect to their intervention logics.	Ignore NRN and TA as RD interventions which need attention in evaluation.
Ensure an iterative approach linking the intervention logic to the evaluation questions, indicators and evaluation methods.	Assess the intervention logic as a one-off element of the evaluation process.

2. EVALUATION QUESTIONS

The purpose of this chapter is to highlight the importance, role and use of evaluation questions in the RDP evaluation, explain their triangular consistency with policy objectives and indicators, to inform about the reduced and revised set of common evaluation questions 2007/2013 and provide guidance for the development of programme-specific evaluation questions.

2.1. The role of evaluation questions in the *ex post* evaluation

Definition of evaluation questions

Evaluation questions are the main instrument for steering and structuring the evaluation. Thus the evaluation questions link the intervention logic of the programme to the CMEF through which programme success is evaluated, conclusions and recommendations formulated and decisions on future policy and interventions are informed.

In this respect evaluation questions (EQs) define the focus of evaluations in relation to rural policy objectives and help to demonstrate the results, impacts, achievements, effectiveness and efficiency of rural development policy. In doing so, they therefore direct the work of the evaluator. The EQs also serve as a basis for defining what kind of information needs to be collected with indicators to ensure that in answering the EQs the evaluator is able to capture and reflect the intended effects of the RDP interventions.

Evaluation questions are linked to cause and effect. They seek to establish attribution: the extent to which the change observed was a result of the RDP interventions. The purpose of evaluation questions is to challenge the intervention logic (**Part II:** Chapter 1 Intervention logic) through observing and analysing changes evidenced by indicators (**Part II**: Chapter 3 Indicators).

Two types of evaluation questions

The Common Evaluation Questions (CEQs) are a set of questions to ensure that the essential impacts of RDPs and achievements towards Community priorities and objectives are analysed across the EU. Common Evaluation Questions ensure that the MA applies a coherent and consistent approach in evaluation. They are linked to the EU rural development policy objectives as stipulated in the Community strategic guidelines for rural development (horizontal evaluation questions) as well as to individual measures. In essence they are related to result and impact indicators. CEQs ensure that the results from national and regional RDPs can be aggregated at EU level.

In summary the **CEQs**:

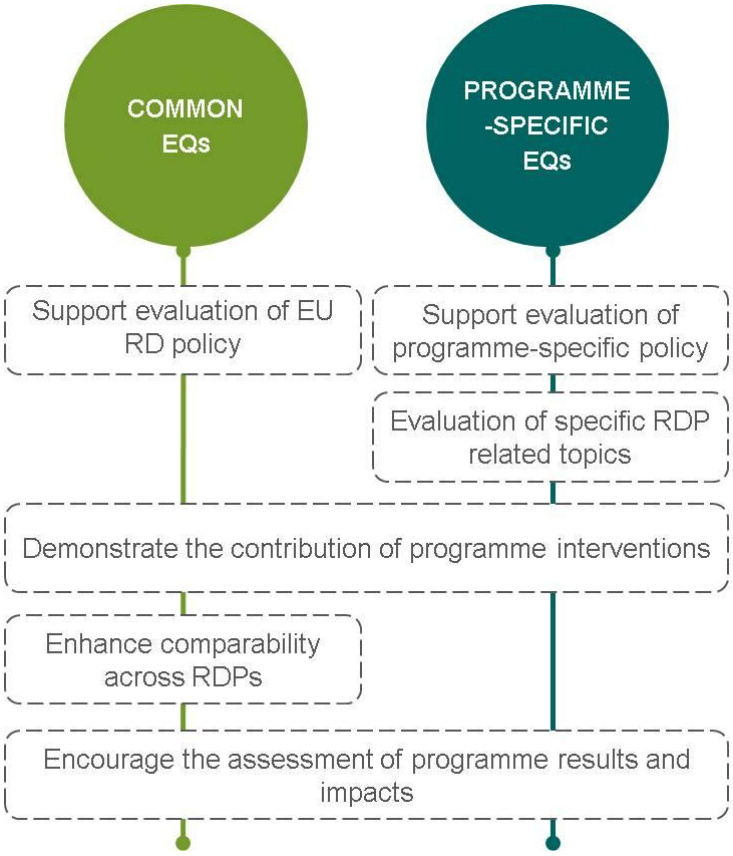
- Steer the evaluation of the EU rural development policies regarding the effects of RDPs interventions towards the EU rural development policy objectives.
- Demonstrate the contribution of EU rural development interventions in addressing the RDP territorial needs.
- Enhance comparability of evaluation results across Europe as part of a commonly applied evaluation system in all Member States or regions.
- Encourage programme bodies and other RD stakeholders to assess results and impacts helping to justify EU policy implementation and support EU policy formulation.

The Programme-Specific Evaluation Questions (PSEQs) are additional to the CEQs and should be formulated for the evaluation of a specific RDP. They help focus the evaluation on the achievement of any programme-specific objectives. These are normally formulated by the RDP Managing Authority to address the specific focus, objectives, implementation or context of the specific RDP or to explore in

greater depth areas thought to be insufficiently covered by the CEQs. As such they form part of the M&E framework of the RDP and therefore should be considered, answered and reported in the *ex post* evaluation in the same way and on the same basis as the CEQs. The programme-specific evaluation questions should be clearly exposed in the terms of reference for the *ex post* evaluations to make sure that these are properly explored by the evaluator.

Figure 7 illustrates the complementary roles and functions of the two main types of EQ.

Figure 7 Two complementary types of evaluation questions



Source: the Helpdesk of the European Evaluation Network for Rural Development

This illustrates that the main roles and functions are common but with two specific differences which contribute to their complementarity: the European-wide comparability on the one hand and RDP specificity on the other.

Linking common evaluation questions, judgement criteria and indicators (triangular consistency)

EQs are answered using judgement criteria and indicators. Judgement criteria developed at the MS level, specify the success of programme interventions and link the EQs, indicators and data to be collected. This enables to design the robust methodological approaches, formulate answers based on qualitative and quantitative evidence in a structured manner, enhance transparency by making the judgement explicit, and improve the objectivity of evaluations.[64]

If EQs are to fulfil their primary role in focusing the evaluation on establishing the link between policy and its effects their use must be demonstrably robust and evidence-based. The links between the EQs, MS specific judgement criteria and indicators used must be logical, consistent and practical; they form part of a system (Figure 8).

[64]EuropeAid Guide to Evaluations

Figure 8 Triangular consistency of EQs, judgement criteria and indicators

Source: the Helpdesk of the European Evaluation Network for Rural Development

A weakness or failure in any one part of this system will compromise the effectiveness of the whole and the value of the other elements however strong as well as it might decrease the quality of evaluation. Since consistency between these elements is critical, all three must therefore carefully be developed and their use planned.

To fine-tune the complete set of EQs (both: common and programme-specific) we recommend thinking through the following:

- The extent to which the CEQ and PSEQ reflect the RDP objectives and intervention logic (including programme-specific effects);
- The extent to which the EQs, judgement criteria and indicators enable the evaluator to capture the full range of achievements from the RDP (including programme-specific effects);
- Consistency with the CMEF and newly proposed set of CEQs to ensure comparability across the EU and allow for sound European synthesis of findings.
- Whether challenges in relation to the use of EQs identified in the MTE, have been addressed;
- Whether EQs can be answered with the envisioned methods;
- Whether sources of evidence and required data are clearly established and accessible and up to date and which fine-tuning is required or would be beneficial.

Provided the evaluation framework meets the essential requirements, a degree of flexibility is introduced in how the evaluation is designed and implemented. The availability and quality of data and information will not be equal for all CEQs across the Member States. Therefore it makes sense to fine-tune the evaluation focus, resources and effort to the specific RDP – under the premise that all CEQs and PSEQs are answered clearly and conclusively.

Evaluation questions in ex post *evaluation report*

According to the CMEF, the *ex post* evaluation is required to contain answers to all (applicable) CEQs and PSEQs derived from an assessment of the effectiveness and achievements, efficiency and relevance of measures and programmes, as well as programme results and impacts.

The indicative template for an evaluation report specifies that in addressing the evaluation questions it should contain:

- Analysis and discussion of indicators with respect to judgement criteria and target levels referred to by evaluation questions;
- Analysis and discussion of quantitative and qualitative data and information from monitoring, public statistics, specific surveys or enquiries, or other sources;
- Answers to evaluation questions,
- Conclusions and recommendations derived from answers to evaluation questions regarding design and implementation of the Rural Development Programme.

2.2. Common evaluation questions proposed for the *ex post* evaluation of 2007-2013 RDPs

The revised set of common evaluation questions

Based on lessons learnt from the mid-term evaluation, the original set of 150 common evaluation questions of CMEF has been simplified and reduced to the essential demand for knowledge from the European perspective. This move serves to strengthen the comparability of the evaluation reports. The aim is also to encourage the Member States or regions to use programme-specific evaluation questions to a greater degree so as to focus on the specific programme in question. The revised and reduced set allows for aggregation of key information at EU level, and provides some information on how outcomes have been achieved. It also gives more room to MAs for the use of programme-specific evaluation questions whenever the common evaluation questions do not address the MS or RDP specific evaluation needs.

It is recognised that this set has limitations, and does not explicitly cover issues such as coherence, delivery mechanisms or gender equality. This was a deliberate choice of questions to be asked at EU level in order to restrict the CEQs to a simple, manageable set which could be accepted by Member States and to which the evaluations can be expected to provide answers. If Member States consider any of these issues, omitted in the CEQ, important for their RDPs, they should include them in the programme-specific questions. For the EU level synthesis of *ex post* evaluations, it is proposed to pick up these omitted themes, using case studies in conjunction with the RDP level evaluations and monitoring data to draw some conclusions on them.

In addition, the common set does not include questions in relation to RDP indirect effects, such as deadweight, displacement, multiplier effects, etc., bearing in mind that these are already included in analyses conducted prior to answering evaluation questions. The common set also does not include the evaluation question in relation to synergies of axes and measures, considering that the synergy issue is already covered by the proposed common set.

There are three groups of CEQs in the revised set (Table 4). The first group of CEQ relates to programme level and reflects the Lisbon treaty, priorities of Community strategic guidelines for rural development[65], CAP Health Check, the 7 common impact indicators, technical assistance, national rural networks and efficiency of RDP implementation. The second group of CEQ relates to measure level and asks for contributions of measure to achievements of axes objectives. The third group of CEQ relates to Axis 4 and asks for contributions of Leader Axis to Community strategic priorities, achievements of rural development objectives and implementation of the Leader approach.

Common impact and result indicators should be employed to answer CEQ. The axis-related common result indicators should be used in the evaluation at two levels: as a contribution to answering the

[65] Community strategic guidelines for rural development , 2006/144/EC

programme-level evaluation questions, in conjunction with the relevant impact indicators, particularly in relation to establishing the net impact of the RDP, and at axis level, where they can demonstrate the achievements of each axis, and contribute to assessing the relative contributions of the measures implemented under each axis.

Table 4 Revised set of CEQs for the *ex post* evaluation.

Programme-related CEQs

1. To what extent has the RDP contributed to the growth of the whole rural economy? *(Lisbon objective; related impact indicators 1:Economic growth and 3:Labour productivity)*
2. To what extent has the RDP contributed to employment creation? (*Lisbon objective*; *related impact indicator 2: Employment creation)*
3. To what extent has the RDP contributed to protect and enhance natural resources and landscape including, biodiversity and HNV farming and forestry? *(Community strategic priority, Biodiversity is also a Health Check objective; related impact indicators 4:Farmland Bird Index, 5: High Nature Value Farming and Forestry and 6: Water quality)*
4. To what extent has the RDP contributed to the supply of renewable energy*? (Health Check objective; related impact indicator 7:increase in production of renewable energy)*
5. To what extent has the RDP contributed to improving the competitiveness of the agricultural and forestry sector? *(Community strategic priority*)
6. To what extent has the RDP accompanied restructuring of the dairy sector? *(Health Check objective)*
7. To what extent has the RDP contributed to climate change mitigation and adaptation? *(Health Check objective)*
8. To what extent has the RDP contributed to improvement of water management (quality, use and quantity)? *(Health Check objective)*
9. To what extent has the RDP contributed to improving the quality of life in rural areas and encouraging diversification of the rural economy? *(Community strategic priority)*
10. To what extent has the RDP contributed to introduction of innovative approaches? *(Health Check objective)*
11. To what extent has the RDP contributed to creation of access to broadband internet (including upgrading) *(Health Check objective)*
12. To what extent has the NRN contributed to RDP objectives?
13. To what extent has the TA contributed to RDP objectives?
14. How efficiently have the resources allocated to the RDP been used in relation to achieving the intended outputs?

Measure-related CEQs

For each of the Axis 1 measures included in the RDP:

15. How and to what extent has the measure contributed to improving the competitiveness of the beneficiaries?

(where relevant, the answers to this CEQ should be presented so that the contribution to the competitiveness of the agriculture and forestry sectors can be seen separately)

For each of the Axis 2 measures included in the RDP:

16. How and to what extent has the measure contributed to improving the environmental situation?

For each of the Axis 3 (Article 52(a)) measures included in the RDP:

17. How and to what extent has the measure contributed to the economic diversification of the beneficiaries?

For each of the Axis 3 (Article 52(b)) measures included in the RDP:

18. How and to what extent has the measure contributed to the improving the quality of life of beneficiaries?

For each of the Axis 3 (Article 52(c) and (d)) measures included in the RDP:

19. To what extent has the measure enhanced beneficiaries' capacities to improve economic diversification and quality of life in rural areas?

For each measure included in Axes 1-3 of the RDP

20. What other effects, including those related to other objectives/axes, are linked to the implementation of this measure (indirect, positive/negative effects on beneficiaries, non-beneficiaries, local level)?

Axis 4 (Leader)-related CEQs

21. To what extent has the RDP contributed to building local capacities for employment and diversification through LEADER? *(Community strategic priority)*
22. To what extent have LAGs contributed to achieving the objectives of the local strategy and the RDP?
23. To what extent has the Leader approach been implemented?
24. To what extent has the implementation of the Leader approach contributed to improving local governance? *(Community strategic priority)*

Table 5 gives an overview on how the evaluation questions, the related judgement criteria and common and programme-specific result and impact indicators are set out in consistency with policy objectives. The evaluator of the individual RDP may fill in the column on the right side with the specific programme-related indicators and additional information.

Table 5 Overview table of policy objectives, CEQs, judgment criteria and indicators

Objectives (programme/horizontal, axis related)	Common Evaluation Question	Judgment criteria	CMEF Indicators	Programme-specific indicators and additional information

2.3. Programme-specific evaluation questions

The role of programme-specific evaluation questions and links to programme-specific objectives

The PSEQ are additional to the CEQs meaning that they should not tackle the issues already embedded in the CEQs. Programme-specific evaluation questions (PSEQ) are usually developed at early stages of programme implementation or prior the mid-term assessment. However they can be defined also for the purpose of the *ex post* evaluation in line with the evolvement of the intervention logic or in case of newly-born evaluation needs.

PSEQs focus the evaluation on the specificities of the RDP, particularly specific elements of its intervention logic, and the achievement of any programme-specific objectives. They bear on programme-specific judgement criteria in assessing the success of these aspects.

Answers to PSEQs are developed with the help of programme-specific indicators. Normally designed by the MA, these questions can also be used to assess elements of their particular interest. More specifically PSEQs:

- Steer the evaluation of programme-specific policies. PSEQs focus the evaluation on programme-specific interventions and their contribution towards programme-specific policy objectives.
- Demonstrate the contribution of programme-specific interventions in addressing the identified specific RDP territorial needs.
- Address evaluation of specific RDP related topics assessing aspects such as programme implementation, management, delivery mechanisms, effectiveness of the communication strategy, etc. which are of particular interest for Managing Authorities.
- Encourage programme bodies and other RD stakeholders to assess results and impacts. PSEQs ask for results and net impacts of programme-specific interventions.

Development of programme-specific evaluation questions

Programme-specific evaluation questions should be phrased in terms of capturing the contribution to the programme-specific objectives sought in terms of the programme results and impacts. As with other evaluation questions these should mainly be cause and effect questions (i.e.: “To what extent did change happen due to the intervention?’’).

The development of PSEQs should follow a similar approach to that for the CEQs and be based on the system and principles identified in the previous section, i.e. linking the questions to policy objectives, specifying them with the judgement criteria and linking them to indicators in a way which demonstrates a strong internal consistency.

The first step in developing PSEQs is for the MA to identify:

- The extent to which the CEQs, judgement criteria and indicators will enable the evaluator to capture the full range of achievements of programme-specific objectives of the particular RDP and the programme-specific effects;
- The extent to which the CEQ reflect the policy objectives and programme-specific elements of the intervention logic.

In case the CEQ are not able to comply with the above mentioned the following steps should be followed:

a. Sketching the PSEQ

In a first attempt, the evaluation question should be drafted with regard to filling the information gap.

b. Development of judgement criteria

A set of judgement criteria should be proposed for each PSEQ; the number and complexity of these will vary with the topic and complexity of the question. The PSEQ related judgement criteria should set the measure by which the success of the programme-specific intervention may be assessed in a given RDP context e.g. the extent to which participation in a supported action has increased.

c. **Formulating the PSEQ**

The PSEQ should now be made definite; it should address the right objective and adequately match the chosen judgement criteria.

d. Linking evaluation questions, judgement criteria and indicators

Once having completed the formulation of PSEQ according to judgement criteria and programme-specific indicators, it is important to check if the established linkages between these elements are likely to provide evidence-based answers.

2.4. How to use the common evaluation questions in evaluation of RDPs

The use of evaluation questions in the preparation of evaluation, structuring, observing, analysing and judging.

In setting up the evaluation system for ongoing evaluation Managing Authorities were expected to discuss the CEQs and develop the PSEQs, specify them with judgement criteria and examine indicators in order to assess what needs to be done in terms of information collection and analysis for answering these questions in a meaningful and appropriate manner. In case this has not been carried out properly, the evaluator will have to "revisit" RDP' intervention logic and (re)formulate appropriate programme-specific evaluation questions and judgement criteria.

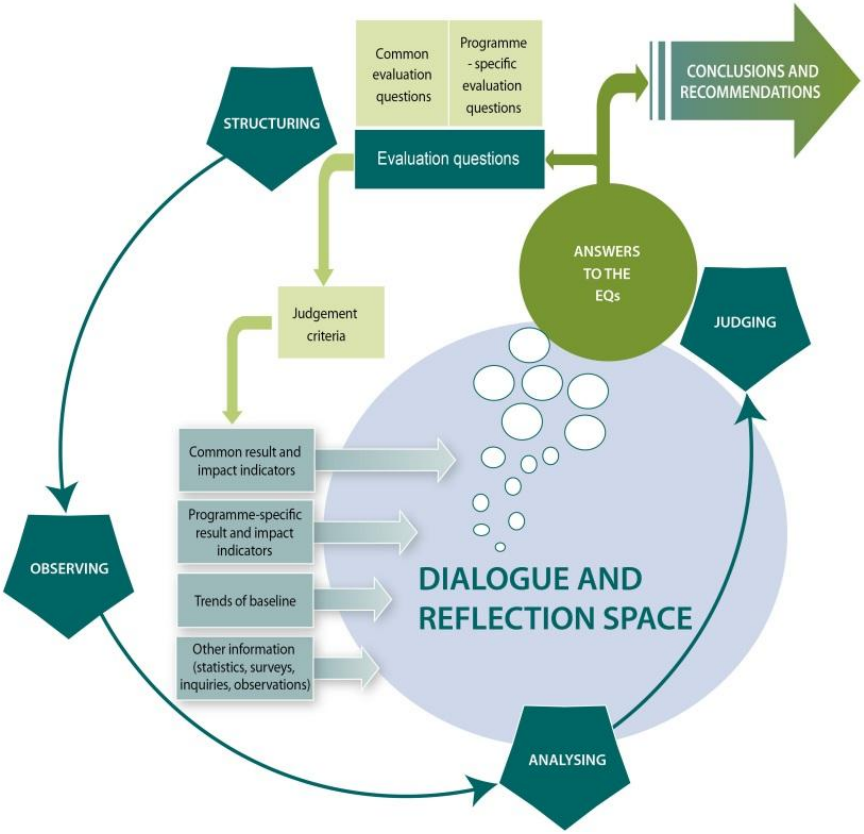
As the EQs are absolutely central to the judgement phase, the key part of the terms of reference for evaluation is the list of common and programme-specific evaluation questions setting out the evaluation themes and referring to the indicators established. It is essential for the MA to provide the evaluator with the resources to be able to work with these effectively, to draw relevant conclusions and develop appropriate and targeted recommendations.

This means that the MA should spend some time on 'retro-planning', working back from the point where the evaluator is involved in the judgement phase and thinking through what needs to happen at each stage to make sure they arrive at this stage with the good quality evidence resources required. This will also inform the structuring of the *ex post* evaluation allowing the evaluator to design tools and implement methods to directly meet the available information needs and further requirements.

Whilst all common and programme-specific EQs should be answered the likelihood is that some will have markedly greater or lesser importance than others. Prioritising EQs or grouping EQs by theme can help to improve their effectiveness and relevance to the evaluation. In setting up and structuring the evaluation the Managing Authority may prioritise those EQs of greatest relevance and importance and where the greatest effort or resource will be concentrated. This is particularly important if resources are limited. In doing so there is a need to consider both practical considerations e.g. data availability and the policy priorities for evaluation, the focus should be on quality and the policy priorities.

The use of EQ in steering, structuring and conducting the evaluation is illustrated with figure 9.

Figure 9 Use of EQs in steering, structure and conducting the evaluation



Source: the Helpdesk of the European Evaluation Network for Rural Development

Answers to the evaluation questions

We are interested only in changes that are directly or indirectly attributable to the RDP. Therefore the answers to the EQs should relate to what has changed *because of* the programme interventions. Answers should be based on the evidence, collected with the means of indicators which are linked to the defined judgement criteria while applying robust evaluation methods. Key terms used in EQs, judgement criteria and indicators should be clearly defined, methods carefully selected. This is also essential to enable comparability and aggregation of CEQs.

PSEQs will provide additional programme-specific information. Where EQs and indicators enable more detailed conclusions to be drawn any such differentiation e.g. by gender, age, region, territory will enable better and more fine-tuned policy recommendations.

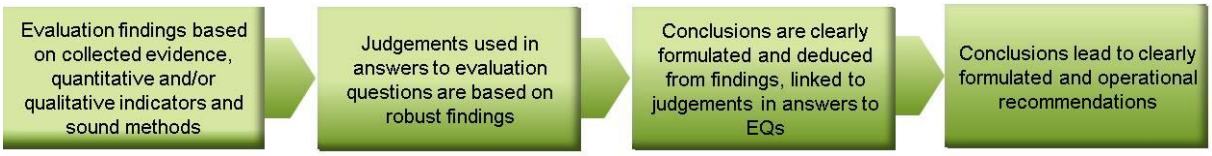
Answers provided to the EQs should indicate the methodology and information sources employed. The soundness of methods and completeness or quality of the evidence used to inform the EQ answers is a key determinant of the validity of the conclusions and recommendations. The answers to the EQs should therefore include a critical appraisal of the evidence upon which they are based. Any constraints, limitations or caveats on the validity of the data or findings should be clearly described.

Reporting

The *ex post* evaluation report will provide answers to all common and programme-specific evaluation questions, derived from judgments developed in the assessment of the effectiveness and achievements, efficiency, and relevance of measures and programmes, as well as programme results and impacts.

The conclusions are based on answers to evaluation questions and should lead to operational recommendations with the view to improve the policy. This relation is illustrated with the Figure 10.

Figure 10 The role of EQs in formulating conclusions and recommendations



Source: the Helpdesk of the European Evaluation Network for Rural Development

2.5. What are the specificities with respect to NRN and TA?

Technical Assistance

The relevant common evaluation question reads: "To what extent has the TA contributed to RDP objectives?"

It is obvious that the exploration of this question requires a hermeneutic approach which is partly based on the monitoring system revealing the quality and completeness of databases, the speed of decision flows, etc. However the judgement will mainly be based on surveys, qualitative interviews, focus groups and similar operations to raise primary information. If feasible, the observation period should stretch over previous programming periods (See **Part II:** Chapter 4 Methods).

National Rural Networks

The revised set of CEQs only marginally covers NRN; the relevant common evaluation question reads: "To what extent has the NRN contributed to RDP objectives?". The common evaluation question might not be sufficient to reflect the NRN specific objectives and NRN specific intervention logic, as well as to capture all results and impacts of NRN, therefore NRN programme-specific evaluation questions need to be developed (e.g. the realm of networking and cooperation, skills acquisition or public dialogue, etc.). Their formulation follows the same basic principles as for all evaluation questions (e.g. link to the intervention logic and the objectives to be achieved, link to judgement criteria and indicators).

For example a NRN defines as one of its objectives: "Strengthening the programming and implementation capabilities in Rural Development Programmes". This objective can be directly transformed into the evaluation question: "To what extent have the programming and implementation capabilities of Managing Authorities and implementation bodies been strengthened through the activities of the National Rural Network?"

Before conducting the analysis to answer this question it is important to clarify several issues:

- the hierarchical level of the objective within the intervention logic: the above objective appears to be placed at the result level; it should be induced by specific activities and outputs, but it still falls short of addressing a policy objective;
- the expected change which objective speaks for: the objective refers to aspired changes in the capabilities in two distinct domains: programming and implementing;
- presumed activities to reach the objective: activities such as trainings, meetings, studies are supposed to make a difference in the capabilities of people, networks and institutions;
- operationalization of key terms: e.g. "capabilities":
    - Capacity building activities: the attendance to trainings, meetings, etc. that can be extracted from the monitoring system;
    - Studies: the thematic selection, commission and reception of studies as well as their follow-up that can be explored by interviewing involved officials and contributing experts;

- Satisfaction of users of capacity building activities that can be assessed by a survey;
- Enhanced interaction among relevant authorities (e.g. between the Managing Authority, the paying agency and the Commission) and other players, to be explored by system environment analysis.
- Other activities…

Once different types of operations have been sorted out, the evaluator can conduct analysis in order to explore the link from distinct operations to observed changes, and also to possible unexpected effects wherever they may surface.

2.6. "Dos and don´ts" in relation to use of evaluation questions (based on experience)

Dos	DON'Ts
Provide answers transparently so that the judgement can be traced back to the chosen method and evidence on which the answers are based.	Provide answers to the EQs without reference to the evidence base and the methods by which the analysis has been conducted.
Develop programme-specific evaluation questions if common set does not address all evaluation needs.	Rely only on common evaluation questions.
Specify the success of the intervention for each of common and programme-specific evaluation questions using the judgment criteria.	Be silent about judgement criteria.
Check the consistency of each of the evaluation questions with judgement criteria and indicators.	Forget about linking programme-specific evaluation questions to indicators.
If inconsistencies and gaps are identified, reconstruct the missing or misleading parts to re-establish consistency.	Overlook consistency gaps between policy objectives, EQs, judgement criteria and indicators.
Apply the proportionality principle and invest relatively more in-depth analysis concerning the EQs which address areas of strategic priority and larger budgetary endowment.	Explore all evaluation questions in the same level of detail, regardless of the importance they have within the whole RDP.

3. INDICATORS

The purpose of this chapter is to recall the centrality of indicators in the evaluation process, to review the requirements related to the CMEF common indicators, to provide guidance for the selection of programme-specific indicators, to consider the use of indicators within the evaluation process, and to expose the reader to the main challenges likely to be encountered in collecting and using appropriate indicators in the *ex post* evaluation of 2007-2013 RDPs.

3.1. CMEF indicators

Definition of indicators

Indicators are one of the key elements through which the achievements of the RDP's objectives are measured.

The essential characteristic of an indicator is that it provides a measurable indication of the performance of a policy, programme, axes, and a group of measures or measure, obtainable at an appropriate level of accuracy. For a complex multifaceted policy such as the RDP, a hierarchy of indicators is required, linked to the hierarchy of objectives, starting with high level indicators connected to the overall impacts, going down to result indicators reflecting results produced from a suite of policy measures and finally coming to immediate output indicators, which measure-specific outputs at the measure level.

Indicators do not stand alone. In order to identify and understand the achievements and outcomes of the policy, indicators have to be substantiated with data, which values need to be analysed by appropriate methods which place them in context.

Types of indicators

For the 2007-2013 programming period, a hierarchy of indicators was established to support the measurement of the impact and achievements of EU rural policies at EU, national and regional level as well as to ensure the comparability of evaluation results across Member States.[66]

Within the CMEF, the following types of indicators have been defined[67]:

- Baseline indicators
- Input indicators
- Output indicators
- Result indicators
- Impact indicators

Baseline indicators: Baseline indicators are those used to assess the starting point of the programme, and provided the basis for the SWOT analysis and definition of the programme strategy. They inform on the conditions (environmental, social and economic) within the territory before the RDP was implemented. Subsequently, they will be used in the *ex post* evaluation to help interpret the impacts achieved by the programme in the light of general trends. They should be used at an appropriate spatial scale which allows the impacts or outcomes of interventions to be assessed. The CMEF identified two categories of baseline indicator, those related directly to the wider objectives of the policy (objective-related) and those linked to general trends likely to have an influence on the

[66] Commission Regulation (EC) No 1974/2006, Annex VIII
[67] CMEF Handbook, Annex 3, Guidance F - K

performance of the programme (context-related)[68]. Example: employment development in primary sector.

Input indicators: These refer to the budget or other resources allocated at each level of the assistance. They provide a reflection of intention, i.e. an allocation of resource to a particular measure or axis of the RDP. They are readily scrutinisable and fully available. Example: expenditure per measure declared to the Commission.

Output indicators: These measure activities implemented within various measures of the programme. The activities are the first step towards realising the operational objectives of the intervention and are measured in physical or monetary units. Linked to individual measures, they provide quantified evidence of expenditure on a particular category of intervention, or the number of beneficiaries of a measure. This information is obtained through the regular monitoring of RDP implementation. Example: number of farms receiving investment support, total volume of investment, area of afforested land.

Result indicators: These measure the direct and immediate effects of the intervention. They provide information on changes in, for example, the behaviour, capacity or performance of direct beneficiaries and are measured in physical or monetary terms. CMEF result indicators are established at axis level, to reflect the fact that interventions under different measures may combine to achieve an observed effect. Experience throughout the programming period has shown that whilst some result indicators can be assessed annually through the RDP monitoring system, others are more suited to periodic assessment by the evaluator. Example: number of farmers participating successfully in training courses, increase in GVA in supported holdings, population benefitting from improved services.

Impact indicators: These refer to the benefits of the programme both at the level of the intervention but also more generally in the programme area. They are linked to the wider objectives of the RDP. Impact indicators are perhaps the most important of all in assessing the success of the RDP. However, they cannot be understood in isolation, and in order to explain observed outcomes, reference may need to be made to output and result indicators. Example: increase in employment in rural areas, increased productivity of agricultural sector, increased production of renewable energy.

Values for CMEF common indicators

The *ex post* evaluation must report on all relevant common CMEF indicators: input indicators – financial expenditure by measure, compared to the initial plan, the common output indicators for each of the measures included in the RDP, all the result and impact indicators, and the provision of updated values for the full suite of baseline indicators.

As described above, the input and output indicator values should be available from programme monitoring data, and should be provided to the evaluator by the Managing Authority. The evaluator will therefore be able to use them, but will not have to devote time and resources to obtaining them.

The data situation for the result indicators is likely to be more varied, depending both on the indicator and on the programme: some result indicators are more readily available than others, and data for some should be provided from the monitoring system. In such cases, validation may be all that is required e.g. result indicators for Axis 2 in relation to land under successful land management linked to specific objectives (biodiversity & HNV, water quality, mitigating climate change, etc.), where the evaluator should check which Axis 2 activities have been counted against each result indicator, verify the intervention logic and check the values. Other result indicators may require significant effort in order to obtain indicator values e.g. increase in GVA on supported holdings, or additional number of tourists. In these cases the monitoring system may go further and contain some relevant data.

[68] For the 2014-2020 programming period, the concept of "baseline indicators" has been clarified: they are all referred to simply as "context indicators", for which the baseline value is the value at the start of the programming period.

Alternatively data may have been collected through ongoing evaluation activities, or the task may be included in the terms of reference for the *ex post* evaluation.

The *ex post* evaluator is likely to be responsible for obtaining indicator values for the seven common impact indicators (and any programme-specific impact indicators), although in some cases some information may be available from ongoing evaluation activities.

Updated values should be provided for all the baseline indicators in the *ex post* report. Many of these will be available from standard statistical sources, at regional, national and/or EU level. Some however, may require additional research or input from the evaluator.

It must be stressed that values for many indicators, in particular result and impact indicators, will require interpretation in the light of overall trends, and/or counterfactuals, in order to draw meaningful conclusions about the contribution of the RDP (See **Part II:** Chapter 4 Methods).

It can be seen therefore, that the focus of the *ex post* evaluator's effort in relation to indicators is likely to be on result and impact indicators.

3.2. Programme-specific indicators

Definition and role of programme-specific indicators

The CMEF made specific provision for RDPs to define additional indicators within the programmes. This provision recognised that the suite of common indicators may not fully capture all effects of programme activity, in particular for national priorities and site-specific measures. The guidance specifies that such additional indicators should be developed by Member States and programme partnerships in a flexible manner, but in accordance with the general principles governing the use of indicators in the CMEF.

The CMEF guidance identifies a number of situations where Member States should provide additional indicators, when:

- common baseline indicator does not cover the specific characteristics of the programme area;
- additional objective or national priority is not covered by common indicators; or
- common indicators do not capture the full benefits of the programme, axis or measure;

A guidance note[69] with examples of additional indicators was provided. Where Member States subsequently developed additional indicators at baseline, output, result or impact level, and included them in the RDP, they become a formal part of the monitoring and evaluation system and must be reported upon in the *ex post* evaluation.

This provision for additional programme-specific indicators was intended to give Member States flexibility in creating a monitoring and evaluation system which was more adapted to their needs and to the scope of the RDP. The guidance made it clear that programme-specific indicators should conform to the principles of the CMEF. In developing additional indicators Member States were therefore required to:

- Ensure the relevance and utility of an additional indicator; i.e. consistency with objectives and evaluation questions.
- Define the type and use of the indicator.
- Ensure that the additional indicator meets accepted quality criteria for the type of indicator and intervention concerned.

[69] CMEF Guidance Note K

It was further recommended to provide a detailed indicator fiche for each additional indicator to facilitate their use in monitoring and evaluation.

Existing programme-specific indicators

The mid-term evaluations (MTEs) reported a rather mixed picture in relation to the use of additional indicators. Nearly half of the MTEs used programme-specific indicators, whilst a further 17% mentioned them, without making significant use of them in the analysis. Only one third did not use programme-specific indicators at all. Whilst some programme-specific indicators were relatively highly developed with clear methods for their use, others were poorly specified, implemented and reported.

Existing programme-specific indicators should be subject to scrutiny of their fitness for purpose in the preparatory phase and structuring stage of the evaluation. Where poorly developed or reported, an assessment of whether remedial action of any sort is feasible should be undertaken. Where this is impractical such indicators should be discarded and the reasons reported.

For example, the *ex ante* evaluator of the 2007-2013 Finnish RDP considered that there was an overemphasis in Axis 3 on economic indicators. Accordingly, "*national indicators have been included in the programme to take into account the community spirit, the bottom-up principle and participation*," The *ex post* evaluator should explore how the issue was addressed in the MTE and whether satisfactory progress has been made in addressing these criteria with appropriate indicators.

Development of new programme-specific indicators

Although it is not required that further programme-specific indicators should be developed in the context of the *ex post* evaluation, the *ex post* evaluator should consider the recommendations of the MTE report in relation to the development of programme-specific indicators. The common CMEF indicators represent a minimal list of compulsory indicators and Members States were expected to add to these. Ideally the final indicator list for the RDP, across all types of indicators, should reflect the chosen emphasis of the RDP, national and EU priorities and objectives.

If significant gaps were identified, and were not addressed during the implementation of the RDP, the *ex post* evaluator may contemplate constructing additional indicators if this is feasible and data is available. The following paragraphs include some factors to be taken into account if the existing set of common indicators is seen as insufficient, and the creation of new programme-specific indicators is under consideration.

Any new indicators should reflect the RDP's objectives, from a Member State and/or regional perspective, as well as they should be consistent with evaluation questions (See **Part II:** Chapter 2 Evaluation questions).

The development of programme-specific indicator is consistent with evaluation question through the judgement criterion, which specifies what should be seen as a success of the intervention in its assessment. There should ideally be a clear distinction made between the criterion and the indicator used, such as is clearly evidenced in the pan-European process of criteria and indicators for sustainable forest management, following the Rio process. Where links between the evaluation question, its judgement criteria and indicators are unclear or incomplete, it may be possible to identify and introduce new indicators or collection of additional information. For example, the CMEF common impact indicator for halted or reduced biodiversity losses is the Farmland Bird Index, which may not adequately reflect all biodiversity changes on farmland and the information on other species must be included in the assessment. As a second example, the desire to reduce the impact of the land-based sector on climate change is assessed by the common impact indicator "Production of renewable energy". This impact indicator cannot capture all the effects of the RDP on climate change mitigation and adaptation.

Inclusion of any new indicators should focus on those which most accurately and appropriately reflect the specific RDP. New impact indicators should connect closely to the overall RDP-specific policy objective, related evaluation question and respective judgement criterion, to return to indicators reflecting climate change, a farm sector GHG budget using IPCC[70] Tier 2 or better would be a welcome addition, because performance with respect to that criterion could be much more accurately reflected if farm level emissions changes were also assessed with an impact indicator.

Whilst most indicators concern the product or outcomes, this may not always be the case. Where particular processes are being promoted such as collaborative actions by farmers or even the Leader approach, there is a case for contemplating the application of process-related indicators, reflecting for example new engagers or recruits to collaborative or participatory actions. The level of detailed information to allow accurate specification of quantitative impact indicators may necessitate recourse to qualitative indicators.

Indicators can be screened against the standard SMART[71] criteria, but we suggest a new mnemonic "RABID" which proposes that what we are seeking is:

- Relevant and representative indicators, and
- Accessible and available indicators, which are…
- Base-lined to the start of the RDP, and
- Indicative of the intervention logic, and…
- Detailed enough to enable attribution at appropriate spatial scale.

In general, the golden rule should be: as few indicators as are needed to answer common and programme-specific evaluation questions adequately reflecting the judgment criteria, so that the evaluation can identify what the programme is trying to achieve, and at the same time allowing seeing the contribution of different measures.

3.3. Experiences in using indicators in the MTE

There is little suggestion that the use of indicators is not a relevant part of the evaluation process. They are recognised as relevant but those undertaking the MTEs experienced enormous challenges in terms of getting hold of values for result indicators, let alone impact indicators. The key lesson from the MTE was that all too often it was impossible to report accurately on many indicators since the collection of good quality indicator data has been very challenging. Some regions or Member States have been rather more effective at implementing the CMEF than others. In addition, there is widespread evidence in the MTE of weaknesses in programme management resulting in poor data entry, fragmentation of data, wrong units used, inappropriate time frames and inaccurate definition (e.g. of FTEs).

As importantly, the indicator needs to be of direct relevance to the measure or programme and change in that indicator should be attributable to the measure or programme. This implies first that the indicator is relevant to the measure and second that the impact of the measure can be separated from other factors that could cause a change in the indicator.

There would seem to be an implicit questioning of the use of targets in the MTE, not least because of the enormous variability of achievement with respect to targets. Where targets were set for result indicators, there was enormous variance around these from +350% of target to -1500% of target, suggesting at least the possibility that the targets were inappropriate. However, it is not reasonable to assume underperformance where targets were not met as there are many confounding variables.

[70] International Panel on Climate Change
[71] SMART - Specific, Measurable, Available, Relevant and Time relevant

Targets can be seen as essential if policy is trying to effect change. Without targets there can be no "shape" to what the programme and measures are trying to achieve..

3.4. How to use indicators in evaluation of RDPs?

A major challenge when using indicators in evaluation of a broad ranging policy instrument such as the RDP which comprises broad-based and multifaceted bundles of policy measures is the attribution of a net policy effect and its separation from other causal or intervening factors affecting the environmental, social and economic realms under scrutiny. We come back to the problem of confounding variables later in this chapter.

Balancing importance and measurability

It is important that indicators allow achievements and effects to be appraised accurately. The accuracy is the extent to which the indicator can be subjected to consistent measurement. On any two days, one cloudy the other bright and sunny, an entomological survey could deliver very different numbers of butterflies in an Axis 2 supported herb-rich meadow. Given that surveys are undertaken at a point in time, it is highly likely that many ecological surveys fail to accurately measure both the number and frequency of species in any particular setting. So default to a Farmland Birds Index may be preferable in terms of its consistent measurability when looking for a biodiversity indicator. There may be many other reasons why impacts in particular are difficult to assess, other than the day-to-day variability in what is being measured. Does the trainee really know how that training experience informed and shaped his subsequent actions and the resultant profitability? Chapter 4 on Methods explores such questions.

Coherence of indicators with the intervention logic, relevance to attribution of effects

The indicators should be coherent with the intervention logic. But the intervention logic can be as much about means as ends. It may specify how a policy target is being achieved and how measures are implemented to ensure maximum chance of success. Ideally, the intervention logic should go beyond identifying the problem and should also justify the means by which the problem is addressed.

It is important to identify whether there is a causal link between an outcome and the application of a policy intervention. What are termed "confounding variables", rather than the intervention itself, may be shaping outcomes. The more multi-causal an outcome, the greater the problem of disaggregating influence of the intervention. The evaluator's task is to select methods that will allow the separation out the influence of confounding variables from impacts of interventions. (See **Part II:** Chapter 1 Intervention logic).

Linking indicators with evaluation questions

The common indicators at each stage in the process (baseline, input, output, result and impact) should connect to the key evaluation questions, which follow directly from coherence with the EU-wide and national policy objectives and measures selected to realise them. The use of programme-specific indicators in this respect is not different from common indicators. They may be necessary to answer evaluation questions where common indicators are not sufficient.

The CMEF evaluation guidelines developed by the European Commission Note B note the obligation on "the Managing Authority to review the common and programme-specific evaluation questions and the related indicators in order to assess what needs to be done in terms of information collection and analysis in order to answer these questions in a meaningful and appropriate manner." However, some Member States or Managing Authorities may not have developed robust approaches to information gathering which meet the *ex post* evaluation needs. Where information is lacking, it may be possible to survey to find that information out. However, this route entails costs, which evaluation budgets may not be able to accommodate. Research projects exploring the problems of estimating impact have often

come up with solutions for a single or a few measures which would consume the entire *ex post* evaluation budget. Compromises are almost always necessary.

Linking indicators and methods

As discussed above, indicators are available measures associated with a judgment criterion that policy makers wish to assess in answering the evaluation question. There is a need to design robust methods enabling the assessment of the extent to which a change in an indicator is attributable to the policy intervention or a range of possible confounding factors.

Where all farmers of a particular class or in a particular land use type (e.g. farmers in areas with natural handicaps) are supported by a measure, benchmarking is difficult and it will be necessary for the method to establish a counterfactual. Where the uptake of a measure is discretionary, there is a need to compare those land managers that take up a measure and those that do not. This creates a substantial challenge in finding comparable holdings and in collecting data at appropriate scale to establish the success or otherwise of a measure. The reference can be constructed from a variety of means depending on whether we are dealing with an economic or environmental measure. The chapter on methods addresses this issue more fully.

Screening data sources and availability

As stated earlier, it is the appropriateness of indicators that makes them important not their ease of measurement or availability ("make the important measurable, not the measurable important!").

Given the obligation to frame the *ex post* evaluation around the structure of the CMEF, and the common and programme-specific indicators, the MA needs to ensure that key information in relation to these indicators is available to the evaluation team. Data required for the evaluation will come from a variety of sources, from programme monitoring, from ongoing evaluation, and from outside statistical sources such as the national statistical office or the national farm accounts data network. Where the Managing Authority has been unable to provide some required data, it may be necessary to include provisions for the *ex post* evaluator to obtain it in the terms of reference (See **Part II:** Chapter 5 Data).

But what would be ideal is often not available. Lack of availability often means that there is not an appropriate available data source that can be used as an indicator and/or that it would entail excessive cost to obtain that data. For example, it is widely acknowledged that farmland birds may be a rather poor proxy for other facets of farmland biodiversity. In the case of support for herb-rich meadows or for specific butterfly populations, there are likely to be more appropriate indicators, but these may not be available in appropriate form or at reasonable cost. So the evaluator is forced to think in terms of "must haves" and "nice to haves"; and, equally the MA must think in the same way when it is monitoring performance.

The level at which data are collected in relation to an objective or an indicator may be inadequate when it comes to assessing the impact of a policy intervention, in that what is discernible as an aggregate problem (say of declining agro-biodiversity) that results in inclusion of particular schemes in the RDP, can only be appraised effectively if there is good survey data of those who are participating and those who are not. This issue constantly arises with respect to discretionary measures. For example, there may be good information about the decline of farmland birds, and even aggregate information about the decline of a particular species. But, unless there is farm-level data of both before and after the policy was introduced and a pairing process is used that explores the bird population changes on a similar farm that does not sign up to that scheme, little reliability can be placed on results. There are so many confounding variables that can influence the numbers of a species in any year, including weather events, conditions in places which the bird uses during migration and wintering and/or breeding season, that attribution to habitat enhancement on an individual farm is usually very difficult to establish with any certainty.

But the collection of information is costly and often recourse to proxy indicators is necessary to avoid excessive cost in policy evaluation. We should recognise that there may be a place for qualitative indicators which describe elements of the programme actions that are especially hard to quantify, such as governance arrangements. It is arguable whether qualitative indicators are indicators awaiting the development of quantification or descriptive components of evaluations that describe the presence or absence and nature of, say, governance arrangements.

In the real world the evaluator is often forced to make compromises. The task should be to conduct the evaluation with an awareness of data weaknesses and to prioritise the tasks in consultation with the MA and EC requirements and resource the evaluation accordingly.

Proxy indicators

There are two types of proxy indicators:

Firstly, in cases where it proves impossible or unrealistic (given time or resource constraints) to obtain sufficient and reliable values for a common CMEF or existing programme-specific indicator, an appropriate proxy indicator is needed. However, given the structure of the CMEF, and the fact that its requirements have been known since the beginning of the programming period, the need for proxies in the *ex post* evaluation should be rare, and suitably justified in each case[72] .

A second category of proxy indicator is where it may be extremely difficult to measure an objective directly. In that respect, a linked parameter considered to reflect the achievement in relation to that objective may be used as an alternative. For example, Axes 3 and 4, both aim to contribute to enhancing quality of life in rural areas. However, “quality of life” is not a clearly defined concept. It is therefore important for the evaluator to become aware of any wide-ranging socio-economic datasets which could be useful. Scotland, for example, has a wide-ranging set of socio-economic data with the most fine-grained level of data availability covering about 500 households. Sometimes such data may yield useful proxy indicators. As an illustration, a change in the number of young adults choosing to stay in an area may be an indicator of perceived quality of life; so too might an increase in the number of farms with a named successor. So, it may be possible to use proxies to assess some dimensions of quality of life and a sense of wellbeing.

Interpretation, working with results, triangulation

As stated previously, even with a comprehensive and appropriate set of indicators, indicator values alone are not sufficient to inform the evaluation. They must be interpreted to exclude the influence of confounding factors, to identify causal linkages, and to avoid misattribution of cause and effect. An effective evaluation seeks not only to identify **what** was achieved, but also **how**, and appropriate analysis is required to answer this question.

For the interpretation of indicator values obtained in the assessment it is very important to triangulate findings. Triangulation of evaluations findings means using several methods in the same assessment to cross check their reliability. This leads to increasing certainty about a result or impact by using two or more methods (usually the combination of quantitative and qualitative) of analysis, often based on different perspectives. For example, qualitative data may be collected from beneficiaries, which throws additional light on findings from quantitative data collected from other sources, including models. Where these two sets of findings agree we can have increased confidence in results. (See **Part II:** Chapter 4 Methods)

[72]Working Document, Defining Proxy Indicators for Rural Development Programmes, Evaluation Helpdesk, Brussels, Draft January 2014

3.5. Specificities with regard to Leader

Axis 4 of the RDP, implemented as Leader, must be covered in the *ex post* evaluation using the same overall principles. However, due to the particular nature of Axis 4, linked especially to 7 principles with which it shall be delivered and the added value which Leader is producing in rural areas compare to "*classical rural development measures in terms of improved governance, generation of social capital, capacity building etc."*, there are some specific considerations which should be taken into account in the evaluation of Leader at RDP level in relation to indicators. In addition the revised set of common evaluation questions (as introduced in the previous Chapter 2) put forward a specific group of CEQs dedicated to Leader. These are:

- To what extent has the RDP contributed to building local capacities for employment and diversification through LEADER? *(Community strategic priority)*
- To what extent have LAGs contributed to achieving the objectives of the local strategy and the RDP?
- To what extent has the Leader approach been implemented?
- To what extent has the implementation of the Leader approach contributed to improving local governance? *(Community strategic priority)*

To answer these questions the CMEF only provides a limited number of common indicators, linked to the Axis 4 measures as output indicators which count the number of local actions groups, total size and population of LAG area, total number of beneficiaries, projects and action supported via LAGs.

As for the contribution of LAGs to the RDP objectives (the 2nd common evaluation question above), the CMEF common result indicators (e.g. gross number of job created, number of participants that successfully ended a training activity) and impact indicators (economic growth, employment creation) can be used as in the case of any other RDP interventions. In case LAGs show contributions which cannot be captured with the common CMEF indicators, RDP-specific indicators can be proposed by the MA for LAGs.

To measure the contributions to local strategy objectives, MA can propose the set of RDP related common indicators. In each case LAGs will develop a set of local strategy specific indicators (mainly result and impact indicators), able to capture its effects on respective LAG territory and achievements towards local strategy objectives.

Regarding the three remaining common evaluation questions, it is obvious that the existing set of common CMEF indicators will not be able to provide sound answers which capture the results and impacts of Leader approach across its full range. In this situation, space is left for the Managing Authority of the RDP to bridge this gap and develop additional indicators.

In order to assist the Managing Authority in developing programme specific indicators for Leader the table in Item 7 of **Part III** provides examples of indicators to be used in the assessment of the application of Leader method (7 principles), for the capturing of added value of Leader (going beyond the contribution to the RDP) with respect to governance and capacity building and answering the common evaluation questions. These examples are neither compulsory nor exhaustive and can be replaced or complemented by other indicators in case the Managing Authority finds better measurement approach respecting specificities of country or region, or develops programme-specific evaluation questions in relation to Axis 4. The MA can also consider proposing these indicators to LAGs and so linking self-assessment and evaluation of Leader.

Regarding the Leader method, indicators are linked to the 7 Leader principles and can be used in answering the common evaluation question:

- "To what extent has the Leader approach been implemented?"

Another set of additional indicators is linked to Leader' s contribution to governance and social capital generation in rural communities[73] as well as to building capacities for employment and diversification and can be used in answering the common evaluation questions:

- To what extent has the implementation of the Leader approach contributed to improving local governance? *(Community strategic priority)*
- To what extent has the RDP contributed to building local capacities for employment and diversification through LEADER? *(Community strategic priority)*

3.6. Specificities with respect to TA, NRN

Technical assistance

Article 66 of Regulation 1698/2005 states that: "*at the initiative of the Member States, for each Rural Development Programme, the EAFRD may finance preparation, management, monitoring, evaluation, information and control activities of programme assistance*". TA is thus about ensuring the smooth design, delivery, monitoring and control and evaluation of the programme.

Technical assistance covers a range of functions from the preparation of the programme right away through to monitoring and control measures. A crucial decision for the MA is the allocation of funds internally or to external experts. Whereas some MSs or MAs may be experienced in delivering complex policy programmes from inception to closure, there may be substantial benefit in using external expertise to address specific issues.

The CMEF contains no common indicators related to TA. However, TA does require assessment during the *ex post* evaluation and where possible it is desirable to stick to the overarching principles of evaluation stated in the guidance documents. We should work from the principles rather than any prescription.

With respect to the hierarchy of indicators, information should be readily available on inputs and outputs i.e. expenditure and simple quantification of the various categories of activity. The output indicators proposed for the 2014-2020 period could provide a good starting point. It is more challenging to identify appropriate result indicators for the evaluation of technical assistance, in particular when this has not been foreseen from the start. The emphasis in the *ex post* evaluation on impacts places the challenge firmly in the hands of the evaluator to design methods that can elicit from a range of stakeholders the perception of the efficiency and effectiveness of the MS's management and delivery of TA, and its contribution to the overall impact of the RDP. The evaluation of TA should be proportionate to the budget.

From an evaluation perspective it is difficult to see much in terms of concrete outcomes or impacts that can be subject to indicator-based analysis. Central to TA, is the ability to assess TA needs and the extent to which those needs were met by either internal capacity or through the externalisation of TA to contractors.

Rather than attempting to address TA with quantitative indicators, we recommend questioning key informants and using Likert scaling to assess the adequacy of TA in meeting the perceived needs, using what might be termed satisfaction indicators. A successful programme would yield evidence from a range of stakeholders that the TA was deemed as quite or very valuable in the main groupings indicated in Article 66. In general, we recommend avoiding a neutral mid-point and pushing the respondent not to "sit on the fence."

We recommend that the extent to which the Health Check generated demands for TA be subjected to scrutiny. We might anticipate that new or increased demands being placed on the MA might result in

[73] Working Paper on Capturing impacts of Leader and of measures to improve the Quality of Life in rural areas, Evaluation Helpdesk, Brussels, 2010

new demands for TA. Particular scrutiny should be placed on the Health Check demands and the extent to which they were addressed effectively, again using a Likert scaling approach.

Where components of TA was externalised we recommend including the external consultants as respondents. This enables a degree of triangulation (not so much on method, as by stakeholder).

National Rural Networks

The CMEF common indicators set[74], while not designed to measure NRN activities, offers several output indicators linked to specific measures which can be seen as relevant to networking activities stipulated in legal framework[75]. These are usually indicators referring to activities like training, participation at events, etc.

Besides of that a lot of space is left for Member States to develop NRN specific output, results and impact indicators when examining the NRN intervention logic and formulating the evaluation questions, reflecting the NRN objectives and action plan.

The most obvious source for output indicators will be the NRN action plan itself, usually containing a basic set of indicators, usually not clearly identified as output- or result-oriented. In most cases, the baseline value is zero.

For example the Maltese NRN action plan foresees activities such as:

- Identify, share and promote good transferable practices;
- Organise exchanges of experience and know-how through seminars, workshops, conferences and focus groups;
- Provide technical assistance for inter-territorial and transnational cooperation, etc.

To examine the achievements of the activities, simple output indicators are provided such as "trainings provided", "newsletters produced", etc. albeit without a link to single activities. At the same time, the Maltese NRN action plan identifies functions that can be translated as "*Specific Objectives*". Some of them are:

- Aid in the set-up and development of the Leader programme in Malta;
- Warrant cooperation and dialogue necessary for bringing about change;
- Development and promotion of common approaches and best practices, etc.

More challenging is the situation in defining result and impact indicators and collecting respective information. The first challenge refers to the baseline situation, since no records usually exist. Hence the change happening during the NRN implementation can theoretically be observed via changing values of these indicators in monitoring (output and results) and ongoing evaluation (results and impacts). However the recording of change is, in the absence of a baseline, subjective and time-sensitive.

Indicators, as well as being well defined and at the same time reflecting the objectives of network programmes, must also be consistent with the evaluation questions at the same level of intervention.

The most challenging part of the evaluation of NRN is the assessment of their impacts using viable and reasonable programme-specific impacts indicators. When examining the NRN, some "generic" network effects can be used as an inspiration for the definition of NRN programme-specific impact (and sometimes result) indicators:

[74] CMEF Handbook, Guidance Notes A, B and C
[75] Council regulation (EC) 1698/2005, Art. 68

- Network effects in relation to the implementation of the rural policy objectives targeted by the RDP itself (related to multilevel sector governance in the narrower sense), mainly those which might foster or weaken the rural development interventions.
- Network effects in relation to the networking capacity of rural stakeholders and the network effects in relation to the achievement of the RDP objectives (related to multilevel sector governance in the broader sense).
- Network effects in relation to the overall governance in the rural areas (related to regional governance proper), i.e. the change in behaviour of rural policy beneficiaries but also rural stakeholders in a broader sense.

3.7. "Dos and don´ts" in relation to the use of indicators (based on experience)

DOs	DON'Ts
Prepare for the use of all CMEF common indicators before the start of the *ex post* evaluation (data collection, etc.).	Wait until the start of the *ex post* evaluation contract to collect data for the CMEF indicators.
Check existing programme-specific indicators.	Use existing programme-specific if not SMART.
Consider the development of additional SMART/RABID programme-specific indicators, if objectives or priorities are not covered by existing indicators.	Use only existing indicators without any critical appraisal.
Develop proxies where necessary.	Use proxies except when justified.
Ensure consistency of indicators with policy objectives, evaluation questions, their judgment criteria and methods.	Use indicators in isolation.
Be creative in seeking data sources and their availability.	Rely only on Managing Authority data provision, or use only data collected by the evaluator, or use low quality of data.
Include information on achievements at LAG level by using indicators defined in LDS.	Rely only on CMEF indicators for Leader.
Look for indicators for TA and NRN.	Ignore the development of programme-specific indicators for TA and NRN.

4. METHODS

4.1 Introduction

An important objective of programme evaluations is the assessment of programme effects[76] and methods play a crucial role in it. Review of world-wide evaluation literature proves that within the last two decades there have been tremendous theoretical and methodological advancements within the field of evaluation. Rapidly expanding research on impact evaluation provided excellent guidance for the development of rigorous evaluation methods which were later successfully applied in the assessment of wide range of policy interventions[77]. Meanwhile there is a wide consensus among social science researchers that *the emergence of systematic policy evaluation and an increasing awareness of the pitfalls associated with inadequate empirical research strategies and methodologies* have been the most important development in the field of applied economics during the last decades[78].

No single method can claim a monopoly for provision of right evidence for policy learning, but the deliberate choice of suitable combination of evaluation techniques (e.g. qualitative, quantitative or mixed), methods (e.g. experimental, quasi-experimental or non-experimental) can lead to robust judgments in the assessment of programme results and impacts and provide solid answers to evaluation questions.

Clearly, applying different methods, even using the same data set, may lead to different results, including the estimated magnitude of programme effects and even its sign (+/-). This fact is however too easily overseen. In consequence, an extensive use of inadequate methodological approaches which are unable to control the selection bias and systematic errors may considerably obstruct an evidence-based policy-making and in extreme cases may lead to deficiently designed programmes or measures being prolonged and/or good programmes or measures discontinued.

Addressing these issues during the evaluation period is not easy and usually is a subject to prohibitive time constraints, as well as capacity and budgetary restrictions. Yet, applying certain criteria and methodologies should help to identify and select the most advantageous and the least biased evaluation approaches[i]. These issues are discussed in the following sections of this chapter.

4.2 Methodological challenges in assessing programme effects in *ex post* evaluation of RDPs

4.2.1 Challenge: distinction between programme results and programme impacts

Evaluators may face difficulties when differentiating between programme results and impacts. This confusion may partly arise from the fact that some parts of evaluation terminology applied by EU differ slightly from those of other international organisations or institutions (e.g. World Bank, Asian Development Bank, USAID, and OECD). For example, these development agencies do not explicitly distinguish programme results from programme impacts and merely include the assessment of programme results under a broader category of impact evaluation[79].

[76]Other objectives of programme evaluations can be: assessment of the need for the programme, assessment of programme design and logic, assessment of how the programme is being implemented, and assessment of the programme's cost and efficiency (see: Rossi, Lipsey and Freeman, 2004).
[77]See: Gertler, et. al; 2011
[78]See: Schmidt, 2007
[79]According to many international organisations (USAID, World Bank, ADB, etc.) "impact evaluation should be based on counterfactual analysis. "Impact evaluation implies a structured test of one or more hypotheses underlying an intervention. Impact evaluations are characterised by a specific evaluation design (quasi-experimental or experimental) in order to answer a cause and effect question. These methods can be used to attribute change at any program or project outcome level, but typically focus on one specific activity. Impact evaluations typically collect and analyse quantitative data, but should also be informed by qualitative data collection methods as long as they are used to gather information from both treatment and

Programme results

Programme results are understood in the terminology of DG AGRI[80] as direct and immediate effects of policy or programme interventions to be calculated or estimated at a programme beneficiary level (causality needs to be verified).

Programme results are measured with result indicators. They can be expected and unexpected, intended or unintended. They provide information on changes, e.g. performance, welfare status or standards (quality of life, environmental, occupational safety, animal welfare, hygiene, etc.) of direct beneficiaries and are measured in physical, monetary or non-monetary terms (e.g. index). Therefore the *ex post* evaluation estimates programme results at the level of programme direct beneficiaries, e.g. farmers, foresters, food processors, micro-enterprises, or small regions (village or communes). It also shows the extent to which a given RD intervention (e.g. measure) affected them. For example, one of results of a RDP (e.g. measure 121) could be an increase of factor productivity in assisted farms, holdings or enterprises. Another would be an improvement of specific welfare or environmental conditions in supported farms, holdings, enterprises or village, etc. Clearly, estimating programme results necessitates the application of evaluation methodologies enabling a separation of effects of a given programme or measure from other factors that may simultaneously influence an observed result indicator at the level of direct programme beneficiary (farms, holdings, enterprises, village or community).

Programme impacts

Programme impacts are in the terminology of DG AGRI causal effects of an intervention lasting in medium or long term (causality has to be verified) and refer to the effect of the programme beyond the immediate direct beneficiaries at the level of the intervention.

Programme impacts are linked to the wider objectives of the programme. According to the terminology used by DG AGRI, impacts should be calculated at programme area level only (e.g. economic growth, change in labour productivity, improvement of water quality, etc.).

Impacts produced by a programme intervention may be positive or negative, primary and secondary, expected or unexpected, intended or unintended. They are normally expressed in "net" terms[81]. , Example: net increase in employment in rural areas, net increase of productivity of the agricultural sector, net increase of production of renewable energy - in the given programme area - due to a given programme. As individual impacts of specific RD measures or a group of measures in a programme area (macro-level) may be negligible and not easily separable from one another, an estimation of programme impact should be carried out for a group of measures (aggregated by Axis) or for a whole programme only (and not for individual measures). Estimating programme impacts usually requires the application of advanced and highly rigorous evaluation methodologies enabling clear separation of programme effects from other confounding factors.

Additionally to calculation of programme results and impacts, an integral part of the evaluation of RDPs should be: i) a better understanding of the processes by which results or impacts are achieved, and ii) to identify the factors that promote or hinder their achievement. Such a feedback, including

comparison groups. <…> Impact evaluations answer cause-and-effect questions about intervention effects <…> and require strong performance monitoring systems to be built around a clear logical framework." see: USAID, 2013.

[80] This terminology varies from those used by the World Bank, OECD, in sense that, according to the latter, programme result is not regarded as a separate effect but it is included in programme impacts. In terminology used by DG-REGIO "result" means a specific dimension of the well-being of people that motivates policy action i.e. that is expected to be modified by the interventions designed and implemented by a policy. Examples are: the mobility in an area; the competence in a given sector of activity, etc. – see: (Glossary) in: "EVALSED: the resource for evaluation of socio-economic development", EC DG-REGIO, September 2013.

[81] "Net" means: after subtracting effects that cannot be attributed to the intervention (e.g. confounding factors,), and by taking into account indirect effects (e.g. displacement, multipliers, deadweight, etc.), see Chapter: "What are programme net effects?"

understanding of mechanisms responsible for adapting successful interventions is a basic component of policy learning.

4.2.2 Challenge: identification of programme effects

Programme effects: The programme effect (result/impact) can be understood as a difference in value of specific outcome[82] (e.g. GVA, labour productivity, value of investments, etc.) for the same unit with the programme and without the programme (under *ceteris paribus* conditions). This definition is valid for any analysed unit of observation (e.g. person, farm, enterprise, community, village, region, programming area, country, etc.) and any outcome which can be plausibly related to the programme. The challenge of evaluation is to assess the programme effects under real world conditions, i.e. to calculate the scope and the magnitude of both intended and unintended changes in specific outcomes that are solely due to policy intervention and not to other factors.

Policy evaluation needs to estimate the causal effect of a given policy or programme on outcomes of interest, on which it is expected to have an impact, i.e. programme effect. Although it is simple to ask evaluation questions about cause and effect, it is not straightforward to answer them and establish whether causal relationships between the programme and observable outcome indicators exist, because other factors may also simultaneously affect a given outcome. For example, observed GVA or labour productivity measure for the group of programme beneficiaries can be affected by a change in input and output prices, managerial skills of beneficiary farmers, factor endowment, etc. The basic issue of evaluation, i.e. what is the effect (result/impact) of a RD programme or measure on a specific outcome (e.g. on the economic performance of a supported unit) constitutes therefore a causal inference problem.

4.2.3 Challenge: Distinguishing the main types of programme effects

Programme direct effects are those affecting programme beneficiaries in an immediate way as a direct consequence of programme support.

In general, any direct programme effect can be grouped into at least three categories: i) direct economic effects, ii) direct environmental effects, and/or iii) direct social effects. Direct effect should always be interpreted as an immediate causal consequence of the programme support on specific economic, environmental and/or social outcomes of interest. Estimating direct programme effects is carried out at the level of direct programme beneficiaries (farmers, enterprises, holdings or communities), requires an assessment of counterfactual in appropriate control group and is carried out together with an estimation of programme results. Here, various result indicators can be utilised as relevant outcomes[83] (e.g. GVA per farm, area under successful land management contributing to the improvement of water quality, gross number of jobs created in a given community supported by the programme, etc.)

Programme indirect effects refer to the programme effects spread throughout the economy, society or environment, beyond the direct beneficiaries of the public intervention.

Indirect effects are usually a derived causal consequence of programme direct effects (economic, environmental and social) and include certain derived effects of an intervention at the micro-level (e.g. deadweight loss, or leverage effects) and macro or regional level. Indirect effects include interactions, externalities as well as general equilibrium effects and they can be intended or unintended. These effects have to be taken into consideration while evaluating overall and net RDP effects. In general, a

[82] An outcome is the state of the target population or the social conditions that a program is expected to have changed (see: Rossi, P, H., Lipsey, W. M., & Freeman, H. E. (2004). Evaluation: A systematic approach (7th ed.). Thousand Oaks, CA: SAGE. Outcomes are defined by means of observable and usually measurable indicators.

[83] Other possible results indicators applied at micro-level can be: profit per holding, profit per ha, profit per person employed, GVA per holding, employment per holding, labour productivity (GVA per person employed), land productivity (GVA per ha), etc. see: Michalek, 2012a.

well-conceived evaluation design should be based on some verifiable hypothesis about whether programme indirect effects are present and why they exist.

A typical feature of programme indirect effects is that they can make an overall economic, environmental and social impact substantially larger than direct effects alone. In the real world, indirect effects occur very often and point out how positive or negative shocks originated from a given programme are transmitted through and/or across local economies. Yet, estimation and especially quantification of various indirect effects may be not an easy task (the outcome depends *inter alia* on the definition of a study area).

In principle, indirect effects may create two challenges for evaluation:

- If a given RDP affects a control group of non-participants this will lead to a biased estimation of programme results or impacts, i.e. the control group will be "contaminated".
- While positive RDP indirect effects may easily exceed programme direct effects, by not accounting for indirect effects the overall programme effects can be highly underestimated. On the other hand, indirect effects can also be very negative. In this sense, failing to have an evaluation design which accounts for presence of indirect effects may result in significant over- or under-estimation of real programme effects.

Programme indirect effects can be grouped in various subcategories, e.g.: induced, dynamic, partial and general equilibrium effects, micro or regional or macro-economic effects, etc.

Programme indirect effects measured at beneficiary level are:

- Leverage effects
- Deadweight effects

Leverage effect is the propensity for public intervention to induce private spending among direct beneficiaries. In cases where a public intervention subsidises private investments, leverage effects are proportional to the amount of private spending induced by the subsidy. Although leverage effects can be estimated directly at the level of programme beneficiaries, their economic and political economy interpretation can be more complex. On one side, subsidisation policies may generate or induce additional consumer spending (e.g. farmers receiving subsidy are also consumers of other goods). However on the other side, leverage effects should be interpreted cautiously and jointly with eventual deadweight effects occurring at beneficiary level.

Deadweight loss effects are those changes observed in economic, environmental or social situation of programme beneficiaries which would even have occurred without the intervention. For example, in case of investment support deadweight loss effect occurs if the RDP participant undertakes the investment also without the RDP support[84].

Programme indirect effects measured at RDP level (programme area) are:

- General equilibrium effects: multiplier effects, displacement effects, substitution effects.
- Other indirect effects.

General equilibrium effects occur when programme positively or negatively affects RDP non-participants. They usually play a more important role in the evaluation of large programmes than in the evaluation of small programmes or RDP individual measures. Evaluation of indirect effects is more complex and requires using advanced methods that can capture these effects. Most important general equilibrium effects are: the multiplier effect, substitution effect and the displacement effect.

[84]The European Court of Auditors report: (decision on grant in relation to the measure implementation with respect to deadweigth) to be found out with Teresa from DG AGRI

- Multiplier effect: is a secondary effect resulting from increased income and consumption generated by the public intervention. Multiplier effects are cumulative and take into account the fact that a part of the income generated is spent again and generates other income, and so on in several successive cycles. In each cycle, the multiplier effect diminishes due to purchases outside the territory. Correctly estimating programme multiplier effects is a rather difficult issue. Among possible methodological approaches, one can find: regional Social Accounting Matrix (SAM), regional I-O models, or general propensity score based models[85].
- Displacement effects are effects obtained in a programme supported area at the expense of other areas. For example, due to a programme support in a specific area, positive employment effects in this area occur at the expense of increasing unemployment in neighbouring areas (or shift of employment from other eligible regions).
- Substitution effects belong to indirect regional effects. Substitution effect is normally defined as the effect obtained in favour of direct programme beneficiaries but at the expense of persons, farms or units that do not qualify or participate in a given intervention. It occurs if, due to support provided from RDP available resources shift (e.g. due to an increase of input prices (costs) or decrease of output prices (profits)), at the detriment of non-supported or non-eligible units (the latter are usually located in close neighbourhood of units directly supported by a given programme).

Other indirect effects e.g. changes in prices or wages within programme areas which may lead to programme substitution effects. Economic shocks such as price increases occurring due to specific RDP measures can be studied within structural or another type of models (e.g. CGE). Building economic models can help to focus on areas where impact of a given programme is larger. However, estimating economic models is not necessarily straightforward.

4.2.4 Challenge: identifying programme net effects

Programme net effects are only those effects which are attributable solely to the intervention (cannot be attributed to other factors beyond the intervention), and take into consideration indirect effects (displacement, substitution, multipliers, etc.).

Programme effects should normally be expressed in “net” terms, which means after subtracting the effects that cannot be attributed to the intervention, and by taking into account indirect effects (displacement, substitution, multipliers, etc.). Sometimes programme “gross” effects are defined as those which are directly observable by the evaluator, e.g. observed change of respective outcome indicator (e.g. income) in the group of direct programme beneficiaries in a given time period (e.g. prior to and after the programme). However, such a classification is misleading because: i) per definition, a programme effect cannot be directly observable; ii) should the above (incorrect) definition be applied, the “net” programme effects would have to be re-defined as those calculated after accounting for a counterfactual situation only. Yet, using the above “net” effects concept would still exclude (not take into account) programme indirect effects (e.g. substitution, displacement, etc.). Given the above and taking into consideration that various definitions of net term exist, the evaluator needs to specify in detail what elements have been accounted for when calculating programme “net” effects.

4.2.5 Challenge: distinction of positive or negative externalities of the RDP and expected or unexpected effects in economic, social, environmental domains

Positive or negative programme externalities are programme effects that may simultaneously occur in various RD domains (e.g. economic, social, and environmental) both at the level of direct programme beneficiaries

[85] See: Becker, et. al., 2012; Michalek, 2012b; ”Counterfactual impact evaluation of EU Rural Development Programmes – Propensity Score Matching methodology applied to selected EU Member States”, Volume 1 – A regional approach, EC, JRC scientific and Policy Reports, http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=5359

and/or non-supported population, or programme area. Negative externalities may also include perverse effects – leading in a precisely opposite direction to that intended (e.g. delaying, postponing or cancelling the investments while waiting for some possible investment support programme; or increased level of bankruptcies of supported enterprises, etc.)

Expected or unexpected effects: Along with analysis of expected effects (as described in the programme logical framework) evaluators should also take note of all types of unexpected effects (positive or negative) occurred as a result of the same programme. These effects may occur at various levels and may constitute an important element of total net programme effects.

While evaluating public interventions one has to think very carefully about the types of programme effects generated by the programme in question. In the majority of cases, effects of a given RDP on its territory are multidimensional, i.e. even a single programme measure (e.g. investment in agricultural holdings) can simultaneously affect various RD domains (e.g. production, income, investment, employment, competitiveness, environment, technical and social infrastructure, etc.) Additionally, many RDP measures can both have intended (usually expected by policy makers) and unintended effects. For example, support of local food processors may lead to negative effects in the form of strengthening local monopolies (e.g. large processors), causing the breakdown of other local food processing businesses, and therefore a decrease in employment and income in non-supported local enterprises, an increase of out-migration, etc.; some investments in irrigation may cause depletion of water resources in other areas, etc.; support provided to a certain type of agricultural producers may have negative effects on non-supported population, etc. *In extremis*, interventions can even have a perverse effect leading in a precisely opposite direction to that intended[86]. In this context capturing the results of socio-economic interventions including unintended consequences and perverse effects is essential. Evaluation should contribute to learning how to better design programmes and how to avoid wasteful interventions[87]. For the reasons above, it is important to ensure that the evaluation of the RDP is comprehensive and includes the assessment of all possible programme impacts and results (positive, negative, expected, unexpected, etc.) occurring in economic, social and environmental domains.

4.2.6 Challenge: identifying factors which may distort the estimation of programme effects

Confounding factors are correlated with exposure to the intervention and causally related to the outcome of interest. Confounding factors are therefore alternate explanations for an observed relationship (possibly spurious) between intervention and outcome. In evaluation one should use methodologies that properly isolate the effect of the intervention from other potentially confounding factors.

Confounding factors are those factors (socio-economic and others) that are correlated (directly or indirectly) with the decision to participate in the programme and outcomes (outcome variable of interest). For example, suppose “better” farmers (e.g. more innovative or with better managerial skills) decide to participate. Such farmers are likely to have higher yields or GVA per ha or GVA per labour unit even without the investment support. A comparison of yields or GVA between these innovative supported farmers and other non-participants is likely to show higher yields and GVA for the supported farmers not only due to programme support but also due to the fact that the supported farmers are innovative and better managers. Effect of confounding factors can be reduced by increasing the types and number of comparisons in the evaluation analysis. For example, to assess an effect of a RDP on GVA in the group of beneficiaries, the initial level of yields (in the period t_0, i.e. prior to the programme)

[86]See: Evalsed, The resource for the evaluation in the socio-economic development, EC, DG REGIO, September 2013

[87]Evaluation literature stresses that an overarching purpose of *ex post* evaluation concerns learning “*through systematic enquiry what public programmes and policies have achieved and understand how they perform in order to better design, implement and deliver future programmes and policies”*. This emphasis on learning underlines a key feature of evaluation that is consistent with the needs of socio-economic development programmes. As these programmes knowledge is imperfect there is constant need to learn about different contexts and how best to combine different measures most effectively (see: EC, DG REGIO, 2013) Furthermore, evaluation has not only to assess which negative and non-anticipated effects occurred, but also to what extent the negative aspects have been appropriately addressed.

could be taken into a list of confounders to ensure that in this aspect (a proxy for "farmers managerial skills" or "innovativeness") the group of programme beneficiaries and controls are more similar. Another example can be the evaluation of environmental measures. Here, the initial environmental conditions should be characterised in a greater detail to ensure that supported and control sites are ecologically similar and therefore less likely to have confounding (i.e. non-programme related) variables influencing observed result or impact indicators.

Selection bias is a special case of confounding factors. It occurs if participants and a control group statistically differ with respect to observable and non-observable characteristics that do have an influence on outcome variables. Not accounting for a selection bias may result in spurious relationship between intervention and outcome, i.e. leading to overestimation or underestimation of true programme effects.

Selection bias in evaluating the effect of an RDP occurs if the mean outcome of those units, which participated in the RDP, differs from the mean outcome of non-supported units even in the absence of support. A reason for this difference could be pre-existing discrepancies in crucial characteristics of both groups which also determine the groups' performance. For instance, agricultural producers may decide whether or not to participate in a RDP based on their former general performance, opportunity costs, expected gains from the programme, etc. Furthermore, RDPs or measures are not assigned randomly to beneficiaries but: i) are designed to target specific beneficiaries with a certain performance characteristic (e.g. under performed producers, enterprises or areas, etc.), or ii) include various eligibility conditions which, in practice, can only be fulfilled by certain types of economic units, e.g. the best enterprises. In both cases, to establish credible counterfactual is problematic: those who do not participate in the RDP are generally an inadequate comparison group for those participating, and it is difficult to tell whether the differences in outcomes of the two groups are due to the differences in the unobservable characteristics or the support from a programme. This makes simple comparisons of both groups' performance statistically biased and therefore unacceptable.

Omission of one or more unobserved variables can lead to biased estimates, therefore an important task of the evaluator is to reduce the selection bias and minimise the role of unobservables by making them "observable" or "controllable" as much as possible.

Self-selection bias is another type of distortion closely related to selection bias. Its extended form occurs if potential beneficiaries (enterprises, farms or communities) that anticipated participation in the programme already adjusted their own performance prior to the start of the programme in order to comply with programme eligibility criteria.

In such a situation, making comparisons of the group of participants with a control group just "before" participation could lead to a significant bias, because the indicators of performance of programme beneficiaries prior to the programme are already influenced by the programme. Self-selection may be based on observed characteristics, unobserved characteristics or both. The important consequence for the *ex post* evaluation is that this type of bias should be recognised and reduced or eliminated (if possible) before the programme impact assessment is undertaken.

Homogeneous treatment effects are programme effects that out of simplicity are assumed to be identical for all programme participants (programme participants would all have the same response to the treatment).

In the case of a homogeneous treatment effect[88] outcomes for programme participants and non-programme participants are two parallel curves only differing in level. Although, such effect would greatly facilitate analysis of programme impacts or results, the assumption of homogeneous treatment effect is inconsistent with empirical evidence and leads to numerous systematic errors and estimation biases (which need to be avoided).

[88] Blundell and Costa Dias (2002) and Caliendo and Hujer (2005)

Heterogeneous treatment effects are programme effects which are assumed to vary across units (beneficiaries and non-beneficiaries).

In case of heterogeneous treatment effects it is assumed that the effect of the support varies across units (a possible effect of an observable component or as a part of the unobservables). Yet, contrary to the homogenous treatment effect, this structure does not allow extrapolation to all population strata of evaluated units. Furthermore, in such a situation (incl. selection on unobservables) it has been proved that the application of the popular econometric ordinary least square (OLS) estimators (even after controlling for observable differences in characteristic) would be inconsistent[89]. In consequence, for heterogeneous treatment effects, other more sophisticated evaluation techniques of impact assessment than OLS have to be applied.

Various biases and systematic errors described above play the role of exaggerating or diminishing programme effects. As they create a threat to the validity of the programme effect estimate an important task of the evaluator is to apply an evaluation design which minimises or reduces them.

4.3 How to design evaluation to address methodological challenges

4.3.1 Using counterfactual in addressing methodological challenges

Counterfactuals: The assessment of programme effects should involve counterfactual analysis. The key in the counterfactual analysis is to construct a group which is as similar as possible (in observable and unobservable dimensions) to those receiving the intervention. This comparison allows for the establishment of causality – attributing observed changes in outcomes to the programme, while removing confounding factors[90].

Conceptually, a basic answer to an evaluation question, "what is an effect of a given RDP on a given unit?" is rather simple: the programme effect is the difference in value of the specific outcome (e.g. GVA, or labour productivity) for the same unit with the programme and without the programme. This definition is valid for any analysed unit of observation (e.g. person, farm, enterprise, community, village, region, programming area, country, etc.) and any outcome which can plausibly be related to the programme.

Ideally, in order to measure programme effects, one would like to design an experiment in which any analysed unit of observation (person, farm, community, etc.) would be exposed to and not exposed to a RDP. Yet, measuring the outcomes for the same unit in two different states (participation vs. not participation) and time is impossible, because at any given moment in time a given unit can either participate or not participate in a given programme. Since a given unit cannot be simultaneously subject to observation in two different states, in any programme evaluation we are always facing a "counterfactual problem". In consequence, in order to measure an effect (i.e. result or impact) of a given programme we have to estimate appropriate counterfactual[91]**.** As we normally observe outcomes for a given unit in a situation of its programme participation only, we have to find a way how to measure what would have happened to the same unit in a situation of its non-participation.

Using the potential outcome model[92], the causal effect of a given RDP on unit i (e.g. farm, region, community, etc.) can be written as:

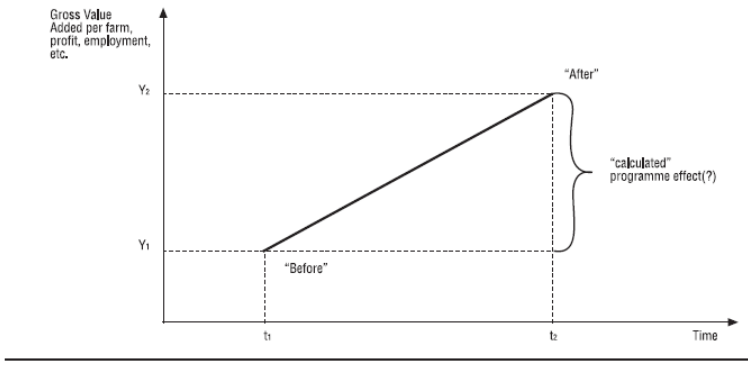
$\tau_i = Y_i(1) - Y_i(0)$ (1)

Where:
$Y_i(1)$ = potential outcome (result/impact indicator) for unit i in case of participation in RDP
$Y_i(0)$ = potential outcome (result/impact indicator) for unit i in case of non participation in RDP

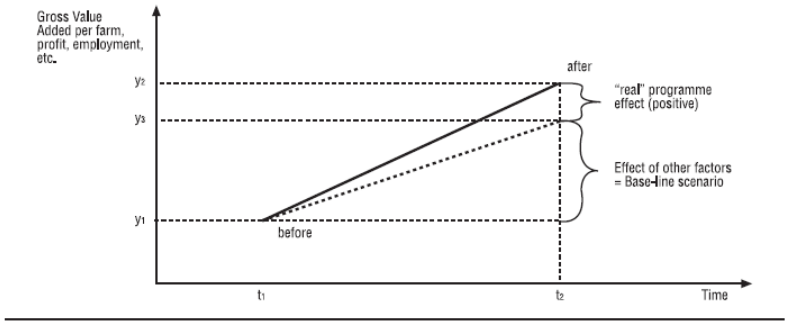
[89] See: Blundell and Costa Dias, 2002
[90] See: World Bank, Impact Evaluation, web.worldbank.org. For definition of confounding factors see: Chapter 4.2.6.
[91] The counterfactual situation is purely hypothetical, thus can never be directly observed. For the same reason, an effect can never be directly observed, nor can an impact (impact indicators notwithstanding). By contrast, effects and impacts can be inferred, as long as the available data allows a credible way to approximate the counterfactual: see: Evalsed Sourcebook: Method and Techniques, EC DG-REGIO, September 2013.
[92] See: Roy, 1951; Rubin, 1974; Holland, 1986

τ_i= the effect of programme participation on unit i, relative to effect of non-participation on the basis of a response variable Y.
In the basic evaluation formula, the term Y_i (0) represents the counterfactual. When conducting evaluation it is relatively easy to obtain for programme beneficiaries the information about $Y_i(1)$ but it is very difficult to estimate Y_i (0) which for programme beneficiaries is not directly observable.
This principle can be illustrated in Graph 1: Estimation of a programme effect (result or impact) for a given observational unit (farm, enterprise, community, region, etc.)

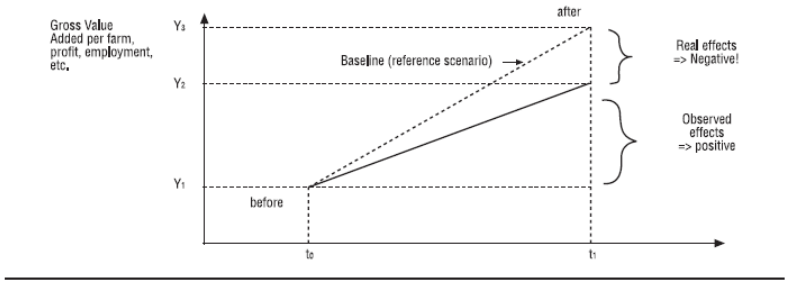
Graph 1a. Naive before-after estimation



Graph 1b. Significance of the relevant base-line ("no programme" scenario) for the same (!) farm/ enterprise (small positive real programme effect)



Graph 1c. Significance of the relevant base-line ("no programme" scenario) for the same (!) farm/ enterprise – negative real programme effect



For a project beneficiary (farm, enterprise, community, region, etc.) one can only directly observe the outcome with the programme starting in period t_1 and ending in t_2, which equals to $(Y_2 – Y_1)$[93]. However, for the same unit the outcome without the programme (counterfactual) cannot be directly observed (in graph 1b and graph 1c) equals to $(Y_3 – Y_1)$[94]). Indeed, we can only observe any observed unit in one of two possible situations: being supported or not-supported which means that the real programme effect (results or impacts)– which in our example equals to $(Y_2 – Y_3)$[95]cannot be directly observed.

The fact that in non-experimental evaluation studies it is impossible to directly observe effect (result or impact) of a given programme does not mean that causal inference is impossible[96]. To the contrary, determining unobservable outcome (called counterfactual outcome) is not only possible but it is generally considered to be the core of any evaluation design[97].

[93]Under naïve methodological approaches (e.g. before-after estimations see: Graph 1a) the whole observed change of a value of a given result indicator (Y2 – Y1) is usually attributed to the programme and is erroneously claimed to be a programme effect
[94]The outcome without the programme (i.e. a change from Y1 to Y3) can be interpreted as a result of other factors which may simultaneously affect observable result/impact indicators.
[95]This is equivalent to: (Y2 – Y1) – (Y3-Y1)
[96]See: Rubin, 1974; 1975
[97]See: World Bank, 2002, 2011; ADB, 2011;

It also means that it is not only an observed change of an outcome indicator for project beneficiary but rather an indirectly observed change of its counterfactual (e.g. a control group) that merely determines the sign (+ or -) and the magnitude of the real programme effect (result or impact).

Evaluating programme effects (also called in evaluation literature "impact evaluation") is essentially about interpreting differences in a causal sense. The challenge facing the evaluator is to avoid giving a causal interpretation to differences that are due to other factors, not to the intervention[98]. Fundamentally, the evaluator always seeks to understand the difference the programme or intervention makes, and behind all evaluations is the counterfactual question: "what would have happened in the absence of the programme or intervention?" Of course, this question can never be answered with certainty because we cannot construct a parallel reality[99]. Yet, the choice of the evaluation design (method and techniques) can be seen as critical in order to answer this key evaluation question and get the most from evaluation resources available.

Methods based on a counterfactual approach became the standard approach to identify the causal policy effects in most institutions and international organisations in the last decades, with the World Bank playing a leading role[100].

Dos and don´ts regarding the construction of counterfactual are summarised in Table 6.

Table 6 The main do's and don'ts regarding the construction of counterfactuals in evaluations

DOs	DON'Ts
Think about the programme effect as a difference between what happens to a given unit (farmer, holding, community, region, programme area, etc.) with programme and what would happen without the programme.	Associate the counterfactual approach with application of any specific evaluation technique or method.
Consider counterfactual as a "normal" logical and necessary component in performing programme evaluation (counterfactual thinking).	Overstretch advanced quantitative methods for finding appropriate counterfactuals. Empirical database may not be rich enough for doing so.
Assign a particular importance to derivation of the valid (the most reliable) counterfactual. This can be done using both qualitative, quantitative or mixed approaches.	Forget that building valid counterfactuals is the most promising technique for the separation of programme effects from other programme independent factors.
Remember that "bad" counterfactuals can be misleading.	Forget that appropriate counterfactuals (control groups) have to be established by an evaluator and not by MA.
Remember that finding valid empirical examples of comparable units which were not affected by a given programme may not always be possible because such control groups may not exist. However, the most advanced evaluation methods do allow finding valid counterfactuals also in situation when all units were supported from a given programme, yet at various intensity levels.	Underestimate the role of a systematic collection of data on programme non-beneficiaries enabling construction of valid counterfactuals.
Remember that counterfactual analysis does not	Hesitate to use also qualitative methods in establishing

[98]See: Evalsed Sourcebook: Method and Techniques, EC DG-REGIO, September 2013.
[99]See: EC, 2013 (ibid)
[100]See: EC, JRC, 2012 (ibid)

DOs	DON'Ts
provide answers to evaluation questions: why? and/or how?	reliable counterfactuals (they can be combined with quantitative methods).

4.3.2 Evaluation designs

There are various ways how to conduct evaluation of policy interventions. They relate to evaluation design, which is determined by the choice of methods used to identify a comparison or control group (counterfactual). Evaluation design can be broadly classified into three categories[101]:

a) Randomised design (sometimes called experimental approach)

b) Quasi-experimental design and

c) Non-experimental design

Randomised (experimental) design

The "Golden standard" in evaluation is randomised controlled experiment or the so-called "experimental design", where randomly selected groups receive support (or "treatment", as technical term) and a randomly selected control group does not. Experimental or randomised designs are generally considered the most robust of the evaluation methodologies. Random assignment ensures that treated and controls are comparable in their characteristics before the policy intervention so any observable differences afterwards can be attributed to the intervention[102.] Random selection from a target population into treatment and control groups is the most effective tool for eliminating selection bias because it removes the possibility of any individual characteristic influencing the selection[103]. In theory, the control groups generated through random assignment serve as a perfect counterfactual, and are free from the troublesome selection bias issues that exist in all evaluations. However, conducting field experiments poses several methodical challenges like external validity, spill-over effects, etc. Moreover, they can also be practically or ethically impossible or socially unacceptable in the evaluation of rural development policies.

Quasi-experimental design

Since it is mostly impossible to randomly assign persons or economic entities to a subsidy or exclude them, other methodological possibilities to design counterfactual have to applied, e.g. quasi-experimental evaluation design. The quasi-experimental evaluation design is very similar to experimental design but it lacks the key ingredient, i.e. random assignment. In quasi-experimental design the standard to base models on individual data is the most effective one and generates results that are much more accurate than those applied at aggregated samples. A crucial issue in an evaluation based on quasi-experimental design is to identify a group of programme participants and a group of programme non-participants that are statistically identical in the absence of the programme. If the two groups are identical (they have the same characteristics), except only that one group participates in the programme and the other does not, then any difference in outcomes must be due to the programme.

Non experimental design

The non experimental evaluation design can be used when it is not possible to identify a suitable control group through application of quasi-experimental methods. Under this design, programme participants are compared with programme non-participants using statistical or qualitative methods to

[101] See: World Bank 2002; 2010; 2011; Asian Development Bank, 2011; Inter-American Development Bank, 1999; EC JRC, 2012; USAID, 2013, IEG (World Bank, IFC, MIGA), 2011b; etc.
[102] Since in experimental design beneficiaries are randomly selected to receive an intervention, each has an equal chance of receiving the programme support.
[103] See: USAID, 2013

account for differences between the two groups.. While this design is relatively straightforward to implement it possesses a number of difficulties[104]. First, the reliability of results is often reduced as the methodology is statistically less robust. Second, the methodology can be statistically complex and may require specialised expertise, especially in the analysis and interpretation of results. Third, although using non experimental design, it is possible to partially correct for selection bias, full correction is very difficult.

In general, applicability of qualitative methodologies to construct a valid counterfactual under non-experimental design is considered as rather limited. Nevertheless, more rigorous qualitative methodologies were recently developed and provide encouraging examples for approaching this problem[105].

An overview of the main advantages and disadvantages of various evaluation designs applied to construction of valid counterfactuals is shown in Table 7.

Table 7 Advantages and disadvantages of various evaluation designs

Evaluation design	Advantages	Disadvantages
Experimental	○ Free from selection bias issues; ○ Ease of measurement (simple econometric methods); ○ Ease of interpreting results; ○ High internal validity.	○ May be expensive and time consuming; ○ Can be politically difficult; ○ Risk of contamination of control group; ○ Difficult to ensure assignment is truly random.
Quasi-experimental	○ Can draw on existing secondary data sources; ○ Can be quicker and cheaper to implement; ○ Evolving econometric methods.	○ Reliability of the results is often reduced, as the methodology may not completely solve the problem of selection bias; ○ Some techniques can be statistically complex that require unique skills.
Non experimental	○ Relatively cheap; ○ Easy to implement since it can draw on existing data sources; ○ Well-developed econometric methods.	○ Reliability of results is reduced as the methodology is less robust statistically; ○ Some techniques can be statistically complex that require unique skills; ○ Full correction of selection bias remains a challenge; ○ Identifying good instrumental variables can be problematic.

Source: Based on Maredia, 2009; and IEG (World Bank, IFC, MIGA), 2011b

4.3.3 Key approaches to evaluation

The range of methodologies that can be applied to evaluation of RDP is very wide. It includes quantitative and qualitative approaches, methods which are better suited to *ex ante* and those which are more applicable to *ex post* evaluations. The main challenge across different evaluation methods is to find good counterfactual – a situation a unit would have experienced had it been not exposed to the programme. Variants of evaluation methods range from naïve approaches, including unverifiable beneficiaries' <subjective > opinion on "experienced" programme effects, simplistic comparisons of the outcomes of programme participants with their pre-programme situations, comparison of outcomes observed in a group of programme beneficiaries to outcomes observed in an arbitrary selected group of non-participants, to rigorous experimental and quasi-experimental approaches, structural and/or econometric models etc. Not surprisingly, in this variety not every methodological evaluation design chosen by the programme evaluator is suitable to correctly answer to evaluation questions. In order to facilitate an appropriate choice, the main approaches to *ex post* evaluation of RDP are briefly presented and their main strengths and weaknesses are discussed.

[104] See: World Bank, 2011
[105] See: Cummings, 2006; Hind, 2010

These approaches can be divided in four categories:

- Theory based approach,
- Quantitative approaches,
- Qualitative approaches (participative approaches belong here we believe),
- Mixed approaches (combination of 3 above) – this should be highlighted that it is the best.

4.3.3.1 Theory-based approach

"Theory of change" which is frequently applied in theory-based evaluations can be described as a way describing the set of assumptions explaining both the mini-steps leading to the long-term goal and the connections between policy or programme activities and outcomes that occur at each step of the way[106].Theory-based approach is an approach in which attention is paid to theories of policy makers, programme managers or other stakeholders, i.e., collections of assumptions, and hypotheses - empirically testable - that are logically linked together[107]. It premises that programmes are based on explicit or implicit theory about how and why a programme will work[ii]. The main characteristic of theory-based evaluation is that it provides an explicit causal chain (or 'theory of change') linking the intervention with specific effects; and then it uses this causal chain to guide the collection of evidence and the analysis of causal contribution by developing hypotheses that can be tested through critical comparisons.

The programme theory explains how programme stakeholders believe a programme will achieve its desired goals. It rests usually upon the stakeholders' experience with how these types of programmes seem to work and also takes into consideration prior evaluation research findings, and more general theoretical and empirical work related to the phenomena under investigation[108.] Chen[109] defines a programme theory as "a specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated."

The most important sources of programme theory are in this context: observations of the programme, exploratory research to test critical assumptions, prior theory and research, and implicit theories of those close to the programme[110].

Empirically, theory-based evaluations seek to test programme theory, to investigate whether, why or how policies or programmes cause intended or observed outcomes. Testing the theories can be done on the basis of existing or new data, both quantitative (experimental and non-experimental) and qualitative. Testing requires gathering of evidence to strengthen the contribution story, using appropriate data gathering techniques, such as surveys and reviewing and analysing administrative data. From a theory-based perspective, several frequently used data gathering techniques can be applied (e.g. key informant interviews, focus groups and workshops, or case studies). Thereafter, newly collected empirical evidence is used to build a more credible contribution story with strengthened conclusions on the causal links in the theory of change.

---

[106]See: Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. Connell, A. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), New approaches to evaluating community 30 initiatives: Volume 1, concepts, methods, and contexts (pp. 65-92). New York, NY: Aspen Institute .

[107]See: Suchman, E. (1967). Evaluative research. New York, NY: Russell Sage Foundation; Chen, H. T., & Rossi, P. H. (1980). The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. Social Forces, 59, 106-122; Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. Connell, A. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), New approaches to evaluating community 30 initiatives: Volume 1, concepts, methods, and contexts (pp. 65-92). New York, NY: Aspen Institute; Pawson, Ray & Nick Tilley (1997). Realistic Evaluation, London; Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. Evaluation, 14, 29-48; Donaldson, S. I. (2007). Program theory-driven evaluation science. New York, NY: Lawrence; EC, DG-REGIO, Theory based evaluation, 2012.

[108]See: Donaldson S.I. and Gooler, L.E.; Theory driven evaluation in action, Evaluation and Program Planning, 26, 2003

[109]Chen, Huey-Tsyh. "Theory driven evaluations", Sage, Newbury Park, 1990

[110]See: Donaldson, Steward, Overcomming our negative reputation: Evaluation becomes known as a helping profession, American Journal of Evaluation 22, 2001, 355-361

Programme theory approach should be applied as a first step to the estimation of the programme indirect effects *inter alia* via providing answers to questions: who might be affected by the programme and how (See **Part II:** Chapter 1 Intervention Logic).

Main advantages

- The basic advantage of this approach is that it helps to explain how and why effects occurred. By following the sequence of stages, it is possible to track the micro-steps that led from programme inputs through to outcomes.
- The importance of theory-based evaluations stems from the fact that a great deal of other information, besides quantifiable causal effect, can be very useful for policy makers to decide on which policy to implement or how to improve the policy design.
- The method may identify a more comprehensive range of effects valued by different stakeholders by engaging them in the process of developing, reviewing and refining the programme theory.
- As for gathering of impacts, the method can help to systematically identify the most needed evidence for the evaluation.
- Concerning applicability of the method to assess causal contribution of a given intervention the approach may encourage systematic use of causal analytical methods to develop and test hypotheses.
- Analysis conducted under a theory-based approach can lead to further data collection to test emerging patterns and hypotheses.

Main disadvantages

- The approach often focuses on stated intended effects only.
- Causal contribution to the intervention is not often explicitly addressed but it is assumed if there is evidence of an expected causal chain.
- Theories underlying a policy or programme are often not directly visible or knowable to the evaluator. They are often not explicitly expressed in official documents. The evaluator has to search for these theories – if they have not been concisely articulated - and to explain them in a testable way.
- Empirical testing of underlying theories can be cumbersome because theory statements may be too general and loosely constructed to allow for clear-cut testing [111], identifying assumptions and theories can be inherently complex.
- The evaluator may have problems in measuring each step unless the right instruments and data are available.
- Problems of interpretation can arise that make it difficult to generalise from results.
- While a specific programme theory may be based on various theoretical and empirical concepts “how the programme works” (including the stakeholders’ former experience), due to its complexity it may fail to be presentable in form of clear functional relationships.
- Due to the complexity of the RDPs full incidence of this specific policy intervention cannot explicitly be examined by checking all possible links as these may be regarded as too multifaceted and not easily summable in their individual properties. In fact, in most cases only impact areas and not the process itself can be empirically observed directly (one of the reason

[111] See: EC, 2012 (ibid).

is the lack of necessary statistical data to be collected continuously, without substantial gaps and at impracticable level of detail).

- Empirical evidences show that overly complex programme theories (especially those linked to various economic, social and political theories) confuse stakeholders and hinder discussions of the important evaluation questions that need to be answered[112].

When to use the approach

It is important to note that theory-based evaluations have to be based on a plausible programme theory established by the intervention logic of RDPs. Thus, the intervention logic, corresponding evaluation questions and indicators guide the consecutive assessment of the individual measures' contribution to achieve its objectives. A theory-based evaluation can be used to test the validity of the assumptions of the RDP intervention logic. The various links in the intervention logic can be analysed using a variety of methods, building up an argument as to whether the theory of the intervention logic has been realised in practice. The method is able to explain why and how results have been achieved and to appraise the contribution of the RDP to the RD objectives. In general, this approach does not produce a quantified estimate of the impact; it produces narrative and non-parametric data such as qualitative classifications (e.g. low, medium, high contribution of a measure to achieve its objectives). The most effective ways to develop a programme theory is therefore an interactive approach involving literature and programme stakeholders[113].

According to EC (2012) impact evaluation guidelines[114], if an experimental setting design is not possible, and if different designs of non-experiments are not possible, then one can move to more qualitative approaches, including a theory of change, to establish counterfactual: "*A theory of change sets out why it is believed that the intervention's activities will lead to a contribution to the intended outcomes; that is, why and how the observed outcomes can be attributed to the intervention. The analysis tests this theory against available logic and evidence on the outcomes observed and the various assumptions behind the theory of change, and examines other influencing factors. It either confirms the postulated theory of change or suggests revisions in the theory where the reality appears otherwise*".

Moreover, theory-based evaluation can contribute to the *ex post* assessment of RD programme effects in situation where impact evaluation, including the counterfactual established through quasi-experimental or non-experimental approach has been carried out, but an explanation of obtained results is lacking.

Preconditions

- Knowledge of alternative theories explaining why and how specific outcomes can be attributed to a given intervention.
- Knowledge of basic testing principles.
- High analytical skills of evaluators.

4.3.3.2 Quantitative approaches

The basic objective of quantitative evaluation methodologies is to enhance the process of measurement of programme effects. There is a broad range of quantitative methods that can be used in evaluation. Yet choosing a particular method in a specific context is not always an easy task for an evaluator, especially because the results may be sensitive to the context, assumptions and methods

[112]See: Fitzpatrick, 2002; Donaldson and Gooler, 2005
[113]See: Donaldson 2001; Donaldson 2003; Donaldson and Gooler, 2005
[114]See: EC, 2012; Theory-based evaluation, (ibid)

applied[115]. In principle, the following groups of quantitative methods can be distinguished for various evaluation designs:

a) Quasi-experimental design:

- Difference in differences (DID)[116]
- The regression discontinuity design (RDD)[117]
- Matching methods, incl. propensity score matching (PSM)[118]
- The combined DID and PSM method (conditional DID)[119]
- Generalised Propensity Score Matching (GPSM)[120]

b) Non experimental design:

- Instrumental Variable (IV) method[121]

[115]See: World Bank, Handbook on Impact Evaluation, 2010, ibid.

[116]Bertrand, Marianne, Esther Dufl o, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" Quarterly Journal of Economics 119 (1): 249–75; Bergemann, A., Fitzenberger, B. and Speckesser, S., (2009). Evaluating the dynamic employment effects of training programmes in East Germany using conditional difference-indifferences. Journal of Applied Econometrics, vol. 24 (5): 797-82; Blundell, R., Costa Dias, M., Meghir, C. and van Reenen, J., (2004). Evaluating the employment impact of a mandatory job search programme. Journal of the European Economic Association, vol. 2 (4): 596-606; Dufl o, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." American Economic Review 91 (4): 795–813.; van de Walle, Dominique, and Ren Mu. 2008. "Rural Roads and Poor Area Development in Vietnam." Policy Research Working Paper 4340, World Bank, Washington, DC

[117]Imbens, G. M., and T. Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." Journal of Econometrics, 142, pp. 615-635; Lalive, R., (2008). How do extended benefits affect unemployment duration? A regression discontinuity approach. Journal of Econometrics, 142 (2): 785-806.; Buddelmeyer, Hielke, and Emmanuel Skoufi as. 2004. "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA." World Bank Policy Research Working Paper 3386, IZA Discussion Paper 827, World Bank, Washington, DC.; Lemieux, Thomas, and Kevin Milligan. 2005. "Incentive Eff ects of Social Assistance: A Regression Discontinuity Approach." NBER Working Paper 10541, National Bureau of Economic Research, Cambridge, MA.;

[118]See: Rosenbaum, Paul, and Donald Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies of Causal Eff ects." Biometrika 70 (1): 41–55.; Rosenbaum, Paul. 2002. Observational Studies. 2nd ed. Springer Series in Statistics. New York: Springer-Verlag; Sianesi, B. (2008). Differential effects of active labour market programmes for the unemployed. Labour Economics, 15(3): 392-421; Caliendo, M., Hujer, R., Thomsen, S., (2005). The employment effects of job creation schemes in Germany. A microeconomic evaluation. Discussion Paper n. 1512, Bonn, IZA. (http://repec.iza.org/dp1512.pdf); Michalek J. (2012), "Counterfactual impact evaluation of EU Rural Development Programmes - Propensity Score Matching methodology applied to selected EU Member States", Volume 1 – A micro-level approach.", European Commission, JRC Scientific and Policy Reports, pp 1-95;
Jalan, Jyotsna, and Martin Ravallion. 2003. "Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching." Journal of Business &Economic Statistics 21 (1): 19–30.; Imbens, G. M., and J. M. Wooldridge. 2008. Recent Developments in the Econometrics of Programme Evaluation. NBER Working Paper No. 14251, National Bureau of Economic Research, Massachusetts; Todd, Petra. 2007. "Evaluating Social Programmes with Endogenous Programme Placement and Selection of the Treated." In Handbook of Development Economics, vol. 4, ed. T. Paul Schultz and John Strauss, 3847–94. Amsterdam: North-Holland; Smith, Jeffrey, and Petra Todd. 2005. "Does Matching Overcome LaLonde's Critique of Non experimental Estimators?" Journal of Econometrics 125 (1–2): 305–53.

[119]Heckman, James J., Ichimura, Hidehiko and Todd, Petra E. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," Review of Economic Studies, Blackwell Publishing, 1997, 64(4), pp. 605-54.; Smith, Jeffrey A. and Todd, Petra E. "Does matching overcome LaLonde's critique of non experimental estimators? " Journal of Econometrics, 2005, 125(1-2), pp. 305-353.; Michalek J. (2012), "Counterfactual impact evaluation of EU rural development programmes - Propensity Score Matching methodology applied to selected EU Member States", Volume 2 – A regional approach", European Commission, JRC Scientific and Policy Reports, pp 1-83;

[120]Hirano, K. and Imbens, G., (2004) The Propensity score with continuous treatment, *Missing data and Bayesian Method in Practice*: Contributions by Donald Rubin Statistical Family; Imai, K. and van Dyk, D.A. "Causal inference with general treatment regimes: Generalizing the propensity score" au *Journal of the American Statistical Association*, 2004, 99, pp. 854-866; Bia, M. and Mattei, A. "Application of the Generalized Propensity Score. Evaluation of public contributions to Piedmont enterprises, al *P.O.L.I.S. department's Working Papers 80*, Department of Public Policy and Public Choice - POLIS, 2007; Michalek J., Ciaian P. and Kancs, d'A. (2014), "Capitalization of CAP Single Payment Scheme into Land Value: Generalized Propensity Score Evidence from the EU", *Land Economics*, May 2014, 90:260-289.; Kluve, J. et.al., "Evaluating continuous training programs using the generalized propensity score, Journal of the Royal Statistical Society: Series A (Statistics in Society) Volume 175, Issue 2, pages 587–617, April 2012; Michalek J. (2012), "Counterfactual impact evaluation of EU Rural Development Programmes - Propensity Score Matching methodology applied to selected EU Member States", *Volume 2 – A regional approach",* European Commission, JRC Scientific and Policy Reports, pp 1-83

[121]Heckman, James J. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32 (3): 441–62.; Heckman, James J., and Edward J. Vytlacil. 2000. "Causal Parameters, Structural Equations, Treatment Effects, and Randomised Evaluations of Social Programmes." University of Chicago, Chicago, IL; Todd, Petra. 2007. "Evaluating Social Programmes with Endogenous Programme Placement and

- Structural modelling and other modelling approaches[122]

c) Naïve estimates of counterfactuals (not recommended for evaluation of RDP)

- Before/After estimator[123]
- "With" vs. "without" approach[124]
- Comparison with population's average

Quasi-experimental design

Under quasi-experimental design a meaningful comparison or control group to match units that received programme support has to be constructed in order to estimate counterfactual and find what would have happened to the programme participants without the programme. A crucial issue is to identify a group of programme participants and a group of programme non-participants (control group) that are statistically identical in the absence of the programme. That means they have the same characteristics or similar characteristics distribution. However it is also important to note that the control group should consist of only those non-beneficiaries who were not affected by the programme in question. In reality, this condition is however not always met (e.g. programme non-beneficiaries might be affected by a given programme in different ways). Should this happen, one may apply various solutions: i) a two-stage approach: first estimate programme indirect effects at a micro-level (e.g. substitution effects) and second, disregard from further analysis those programme non-beneficiaries who were found to be indirectly affected by the programme in question[125]; ii) include in the control group only similar units (programme non-beneficiaries) that were located in regions where the programme (or a given measure) was not implemented; iii) when all units were supported from a given programme, apply more advanced techniques (e.g. generalised propensity score matching) to all regions, communities or enterprises which received programme support - yet, at various intensity level (in this case respective controls can consist of programme beneficiaries). In such a case any difference between groups must be due to the programme.

Based on the above, the establishment of a similarity between supported and control groups involves the following criteria:

- The supported group and the control group should be identical in the absence of the programme;
- The supported and control groups should react to the programme in the same way;
- The control group should not be affected by the programme in question; and
- The supported and control groups cannot be differentially exposed to other interventions during the evaluation period.

Selection of the Treated." In *Handbook of Development Economics*, vol. 4, ed. T. Paul Schultz and John Strauss, 3847–94. Amsterdam: North-Holland; Wooldridge, Jeffrey. 2001. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

[122]Essama-Nssah, Boniface. 2005. "Simulating the Poverty Impact of Macroeconomic Shocks and Policies." Policy Research Working Paper 3788, World Bank, Washington, DC.; Bourguignon, François, and Francisco H. G. Ferreira. 2003. "*Ex Ante* Evaluation of Policy Reforms Using Behavioral Models." In *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*, ed. François Bourguignon and Luiz A. Pereira da Silva, 123–41. Washington, DC: World Bank and Oxford University Press; Treyz, F. and Treyz, G. " Evaluating the Regional Economic Effects of Structural Funds Programmes Using the REMI Policy Insight Model", *Regional Economic Models, Inc. Challenges for evaluation in an Enlarged Europe,* Budapest, June 26-27, 2003, , 2003.; Rickman, D. S. and Schwer, R. K. "A Comparison of the Multipliers of IMPLAN, REMI, and RIMS II: Benchmarking Ready-Made Models for Comparison." *The Annals of Regional Science, 1995,* 29(4), pp. 363-374.

[123]For critical analysis of this approach see: World Bank, 2010, 2011; ADB, 2011;

[124]See Footnote 145.

[125]For details see: Michalek, 2012a" Counterfactual impact evaluation of EU Rural Development Programmes – propensity Score Matching methodology applied to selected EU Member States", Volume 1–micro-level approach, EC, JRC Scientific and Policy Reports.

If these conditions are met, confounding factors, which may also affect result indicators, are removed, so causality of the programme effects can be ensured.

In principle, programme effects can be calculated using three specific indicators (see: Box below):

- Average Treatment Effect (ATE) measuring the effect of a programme on the population of programme beneficiaries and non-beneficiaries (randomly selected).
- Average Treatment on the Treated (ATT) – measuring the effect of a programme on programme beneficiaries.
- Average Treatment on the Non-treated (ATNT) – measuring the effect of a programme on programme non-beneficiaries.

A treatment effect is the calculated causal effect of a programme on an outcome variable of interest, e.g. GVA per farm, for selected units of analysis (farmers, enterprises, holdings, communities, etc.).

In empirical applications assessing programme direct effects is usually carried out on the basis of the ATT (Average Treatment of the Treated) indicator using data from "before" and "after" implementation of the programme.

Average Treatment Effect (ATE)

The first treatment effect indicator which can be applied to evaluation of RDP is the (population) average treatment effect (ATE). This indicator is simply the difference between the expected outcomes after participation in the RDP and non-participation conditional on X[126].

$\Delta^{ATE}(x) = E(\Delta|X = x)$, where: (2)

$\Delta = Y_1 - Y_0$

X = vector of specific characteristics for all units **i**

Y_0 = effect in case of programme non-participation

Y_1 = effect in case of programme participation, and

E(.) denoting the expectation operator

A characteristic property of ATE indicator is that it describes effect of assigning programme participation randomly to every unit i of type X (ignoring programme general equilibrium effects). In evaluation ATE describes an expected gain from participating in the RDP for a randomly selected farm, individual, enterprise, sub-region, etc. from the joined sub-groups of programme participants and non-participants in a given territory (e.g. programme area). This policy indicator averages the effect of the programme over all units in the population, including both programme participants and non-participants. As every specific policy indicator ATE has also some disadvantages, of which the most important concerns an inadequate addressing of the targeting of intervention. Here, problems may arise due to the fact that ATE includes the effect on units, farms, individuals, etc. for which the programme was never intended or designed (it may include the impact on units that may even be programme ineligible). The ATE effect is a weighted effect of ATT and ATNT (see below).

Average Treatment on the Treated (ATT)

Given deficiencies of ATE, another evaluation indicator was developed to analyse the average effect of programme participation on programme beneficiaries only, i.e. on units, farms or individuals that participated in the programme, the so-called Average Treatment on the Treated (ATT). As most evaluation questions are focused on the effect of a programme on direct programme beneficiaries ATT indicator is the most popular indicator to estimate programme effects. The average treatment on the treated (ATT) can be expressed as in eq. 3:

$\Delta^{ATT}(x) = E(\Delta|X=x, D=1)$ [127] (3)

Where:

D =1 programme participation

In contrast to ATE, interpretation of ATT is much more policy relevant. While ATT focuses on the effect of the

[126] See: Heckman, 1996; Imbens, 2003; Imbens and Wooldridge, 2007

[127] ATT is equivalent to: $E(Y_1-Y_0|D=1) = E(Y_1|D=1) - E(Y_0|D=1)$, where: $E(Y_0|D=1)$ is not directly observable and has to be estimated as counterfactual.

programme on programme participants only, it also describes the gross gain accruing to the economy from the existence of a given programme compared to an alternative of shutting it down[128]. Combined with information on programme costs and general equilibrium effects the ATT indicator can answer the policy question regarding the net gain to the economy. Although ATT is generally applicable to provide answers to common evaluation questions concerning the effect of the RDP on units that participated in the programme, the empirical estimation of ATT is not straightforward because it requires estimating appropriate counterfactuals[129]. In practice, ATT estimators are widely applied in programme evaluations to assess programme *direct effects*. Beyond this, ATT indicators can also be applied for assessment on programme *indirect effects.*

Average Treatment on the Non-Treated (ATNT)

Measuring the RDP effects on those who did not participate in it can be of considerable interest to the evaluator.

The ATNT effect is defined as:

$ATNT = E(Y1| D = 0) – E(Y0| D = 0)$ (4)

The ATNT is the average response to programme for a sample of individuals, farmers, enterprises or communities that chose or had **not** been assigned any programme support. As E (Y1|D=0) (in eq 4) cannot be observed directly it must be calculated as counterfactual.

In order to estimate programme impacts, direct programme effects calculated at beneficiary level should be *aggregated* (to a programme level) and thereafter contrasted with aggregated programme indirect effects in order to compute net impacts.

Main advantages:

- Quasi-experimental methods share similarities with the experimental design or randomised control trials and therefore are closer to the « golden standard » in evaluations.
- They are exceptionally useful in areas where it is not feasible or desirable to conduct an experiment or randomised control trial.
- They are easier to set up than true experimental designs.
- They typically allow the researcher to control for other factors not related to policy intervention and therefore enable to draw conclusions about causal inferences between the programme and outcome indicators.
- They allow for generalisation to be made about the total population.
- They contribute to a greater accuracy of the results by enabling to partial out effects of confounding factors.
- They are very suitable for measuring the effectiveness of a given programme (including comparison with target values) and serve as a valuable input to cost-effectiveness analysis.
- They provide necessary and valuable information for applied evaluation research although on their own they do not allow making definitive causal inferences (they cannot control for all observed and unobserved heterogeneities).

Main disadvantages:

- Reliability of the results is often reduced as these methods cannot completely solve the problem of selection bias (confounding factors may also include unobservable factors which cannot be easily measured and controlled).

[128] See: Heckman and Robb, 1985; Heckman, 1997; Smith, 2000; Smith and Todd, 2005

[129] To illustrate the problem we consider both components of ATT (i.e. E (Y_1| D =1) and E (Y_0|D =1). It is obvious that E (Y_1| D =1) can be easily identified from data on programme participants. In practical evaluations, the term E (Y_1| D =1) describes specific outcomes (e.g. in form of result indicators), e.g. profits, employment, labour productivity or total productivity, etc. observable among programme beneficiaries after implementation of the given RD programme. On the other hand, the expected value of (Y_0|D =1), i.e. the counterfactual mean in outcome (potential outcome in case of non-participation) of those who participated in the programme cannot be directly observed. Given the above, one has to choose a proper substitute for unobservable E(Y_0| D = 1) in order to estimate ATT.

- Some specific quasi-experimental methods and techniques can be statistically complex and require high quantitative skills (e.g. advanced econometric knowledge).
- Some specific quasi-experimental methods (e.g. propensity score matching) are "data hungry".

When to use the approach:

In principle, quasi-experimental approach should be used as a standard evaluation design aiming at quantitative assessment of effects of RDP. However, as no single evaluation method is perfect, verification of quantitative findings should be done by combining quantitative and qualitative approaches (a mixed approach is recommended). Evaluation practice shows that application of particular quantitative method under quasi-experimental design can be problematic in case of insufficient data.

Preconditions:

- Abundant data (general characteristics and performance) of programme beneficiaries and non-beneficiaries prior and after implementation of the RDP. In case there are no non-supported farms, communities or regions abundant data on supported units is necessary (in this situation an application of GPSM method – see below - should be considered).
- Evaluators with high quantitative skills and knowledge of advanced statistical techniques.

Techniques used in the quasi-experimental design

Technique	Double-difference (DD) or Difference in Differences (DID)
Description	The technique compares the before/after changes of programme participants with the before and after changes of outcome indicators for arbitrary selected non-participants. Estimating programme effects is carried out in two steps: i) calculating the average difference in outcomes separately for programme participants and non-participants; and ii) by taking an average difference between the average changes in outcomes for these two groups. The difference in the before and after outcomes for the programme supported beneficiaries (the first difference) controls for factors that are constant over time in that group. After subtracting the second difference (in control group), we may eliminate the main source of bias that was present in the simple before and after comparisons. Under the DID method, the group of programme beneficiaries has no need to have the same pre-support conditions as the group of non-beneficiaries. But for DID to be valid, the group of programme non-beneficiaries must accurately represent the change of outcomes that programme beneficiaries would have experienced in the absence of the programme. This is far from being self-evident. The difference-in-differences approach thus combines two "naïve" techniques, i.e. before and after comparisons of programme beneficiaries, and comparisons between programme beneficiaries with programme non-beneficiaries, to produce a better estimate of the counterfactual. A standard DID method is therefore more advanced compared to naïve techniques (the latter are not recommended for evaluation of RDP) as it additionally assumes that selecting a programme depends on both observables as well as unobservables. Although in this method any common trend in the outcomes of programme participants and non-participants (fixed selection bias) gets differenced out, the crucial assumption justifying this method is that selection bias remains constant over time (so-called fixed effect). In consequence, the DID method cannot help to eliminate differences between programme beneficiaries and non-beneficiaries that change over time.
Advantages	○ The method allows for control of unobserved heterogeneity (which may have an impact

	on outcomes), yet only under the assumption that this does not vary in time. ○ Flexible approach allowing for illustrative interpretation. ○ Can be easily combined with matching estimators to allow for better accuracy of results.
Disadvantages	○ The method assumes that in the absence of the programme, the performance of beneficiary group would move along the same trend as in the group of non-beneficiaries (although the latter group is NOT matched). If however, in the absence of the programme, outcome trends are different for the group of beneficiaries and comparison groups, then the estimated treatment effect obtained by difference-in-difference methods would be invalid, or biased. ○ The method is generally less robust than other quasi-experimental methods (e.g. PSM or GPSM). The common trend assumption might not be verified or might not be testable.
When to use the technique	The approach can be used under the condition that without the programme, outcomes would increase or decrease at the same rate in both groups – a good validity check whether this assumption holds or not is to compare changes in outcomes in both groups in a longer time period before the programme was implemented. Overall, the available evidence shows that standard DID estimators, though motivated by plausible stories about "fixed" differences in motivation, ability or performance may be not sufficient choice in many evaluation contexts.
Preconditions	○ The method requires either longitudinal or repeated cross-sectional data (time series of cross-sectional data) on outcome indicators collected for programme beneficiaries and non-beneficiaries. ○ Strong evidence that in the absence of the programme the economic performance of programme beneficiaries and non-beneficiaries will be the same (similar). ○ The support must have occurred between the two periods observed by researcher.

Technique	The regression discontinuity design (RDD)
Description	The technique can be used to assess the effects of programmes or measures that have a continuous eligibility index with a clearly defined cut-off score determining which farms, enterprises, holdings or communities are eligible and which are not. The main idea behind this design is that units in the target population just below the cut-off (not receiving the intervention) are good comparisons to those just above the cut-off (exposed to the intervention). Thus, in this setting, the analyst can evaluate the impact of an intervention by comparing the average outcomes for the recipients just above the cut-off with non-recipients just below it. Under certain comparability conditions, the assignment near the cut-off can be seen almost as random. The RDD method assumes that individual units around the eligibility cut-off point (on both sides) are similar; thus the selection bias should be minimal.
Advantages	○ RDD can be used for programmes that have a continuous eligibility index with clearly defined cut-off score to determine who is eligible and who is not. ○ In the neighbourhood of a cut-off for selection a RDD is very similar to an experiment ○ RDD method allows to identify the programme's causal effect without imposing arbitrary assumptions on the selection process (without excluding any eligible population), functional forms, or distributional assumptions on errors.
Disadvantages	○ The estimated impact is only valid in the neighbourhood around the eligibility cut-off score. There is no comparison group for units located far from the cut-off point. ○ While applying this method very little can be said about the effect of the programme on units located far away from the cut-off point (in each direction). ○ Similarity between individual units is defined here on the basis of eligibility criteria only, and does not take into consideration other observable characteristics which can also be important to explain a selection bias. ○ RDD results may be sensitive to the functional form used in modelling of the relationship between the eligibility score and the outcome of interest. Moving further from the cut-off may therefore require additional functional form of assumptions to obtain a credible estimate of the impact. ○ The RDD method allows for estimating local average effects only, around the eligibility cut-off at the point where supported units comparison units are the most similar. ○ In practice, a determination of a bandwidth around a cut-off score is rather arbitrary. ○ The RDD method cannot be applied to estimate an average treatment effect (ATT) for all programme beneficiaries. ○ External validity of this method is very limited (it produces local average treatment effects that cannot be generalised).
When to use the technique	RDD can be used when the programme foresees a distinctive eligibility threshold (who is eligible and who is not) on the basis of one or more criteria. To apply a regression discontinuity design, two main conditions are needed: ○ A continuous eligibility index on which the population of interest can be ranked, and ○ A clearly defined cut-off score, that is, a point on the index above or below in which the population is classified as eligible for the RD programme or measure.

Technique	The regression discontinuity design (RDD)
	Depending on the nature of the eligibility rule (whether it changed over time), panel or cross-sectional data should be used. The main steps in implementing RDD methods should include: a) Specifying the selection variable and the threshold, b) Choosing the interval around the threshold; c) Checking the type of discontinuity; d) Estimating programme effects; e) Discussing the external validity of obtained results.
Preconditions	○ Available dataset containing the selection variable and observations on eligible and non-eligible units. ○ Time series of cross-sectional data.

Technique	Matching methods, incl. propensity score matching (PSM)
Description	The technique is currently one of the most advanced and effective tools applied in evaluation of various programmes, especially if combined with DID method (see below). One can distinguish two types of PSM approaches: i) standard, conventional or binary PSM; and ii) Generalised PSM. The PSM is a powerful quasi-experimental approach which can be used to find appropriate controls using counterfactuals and estimate the programme effects in a relatively straightforward manner. The degree of similarities between different units is measured on the basis of the probability of being exposed to the programme, given a set of observable characteristics (not affected by the programme), the so called propensity score. The idea is to find, in the case of binary PSM, from a group of non-participants, units that are observationally similar to programme participants in terms of pre-programme characteristics. Participants are then matched on the basis of this probability (propensity score) to non-participants. Each participant is matched with an observationally similar non-participant, and then the average difference in outcomes across the two groups is compared to calculate the programme treatment effect. In practice, different techniques or algorithms can be used to match participants and non-participants on the basis of the propensity score. They include the nearest neighbour (NN) matching, caliper and radius matching, stratification and interval matching, kernel matching, or local linear matching (LLM). Implementing a binary PSM to assess programme effects consists of 5 steps: a) econometric estimation of propensity scores for supported and non-supported units; b) checking overlap and common support region; c) choosing the best matching algorithm; d) performing balancing tests (checking matching quality); e) estimation of programme effect (e.g. using Average Treatment of Treated indicator).
Advantages	○ PSM is an effective tool applied in impact evaluation of various programmes. ○ PSM is an effective tool to find appropriate control groups (counterfactuals). ○ PSM works well under the assumption that programme participation is affected by the unit's (farm, person, community, region, etc.) observable characteristics only. ○ If a selection bias from unobserved characteristics is likely to be negligible, then PSM provides a good comparison with randomised estimates. ○ PSM is a semi-parametric method and very flexible approach, imposing fewer constraints on the functional form of the treatment model, as well as fewer assumptions about the distribution of the error term.

Technique	Matching methods, incl. propensity score matching (PSM)
	○ Binary PSM does not require any functional form of assumptions for a relationship linking the outcome variable with the covariates. ○ PSM allows to compute ATT (Average Treatment Effect on Treated), ATE (Average Treatment Effect), ATNT (Average Treatment Effect on Non-Treated) indices straightforwardly.
Disadvantages	○ A strong assumption behind PSM is that the observable characteristics used to estimate the propensity score, explain all differences between the supported units and the comparison group prior to programme implementation. ○ PSM works under the assumption that there are no systematic differences in unobserved characteristics between the units supported by the programme and the matched comparison units that could influence the outcome. ○ The PSM method will not generate reasonable results if there are other important observable characteristics explaining the differences in performance but that were not included in the model, or if there are unobservable characteristics explaining these differences but which are not constant over time. ○ PSM requires a relatively large amount of data about supported units and comparison groups before the programme (if PSM is applied together with DID also data after the programme) ○ The effectiveness of PSM also depends on having a large number of participant and non-participant observations in order to find a substantial region of common support. ○ The outcome of PSM is only true (external validity is large) if the region of common support is large. ○ PSM is a data-hungry procedure: its conclusions only hold on the subset of matched units. ○ The external validity (generalisability) of its results decreases when the share of unmatched units increases.
When to use the technique	While applying a binary PSM method to identify a valid control group, one has to be sure that there are no systematic differences in unobserved characteristics between the units supported by the programme and the matched comparison units that could influence the outcome. Obviously, this method will not generate reasonable results if there are other important observable characteristics explaining these differences but that were not included in the model, or if there are unobservable characteristics explaining these differences but that are not constant over time (yet, various observable proxy variables can be used as controls for unobservables; when unobservable characteristics are fixed over a period of programme implementation, the effect of these unobservables can be controlled by combining the PSM with the DID method – see below).
Preconditions	○ Abundant data on programme participants and non-participants. ○ High quantitative skills of the evaluator. ○ Good understanding of conditions determining probability of the participation in the programme.

Technique	The combined PSM and DID method
Description	The technique (also called conditional DID estimator) is a highly applicable estimator when the outcome data on programme participants and non-participants is available both "before" and "after" periods. PSM-DID measures the effect of the RDP by using the differences between comparable to programme participants and non-participants in the before/after situations. In this method observed changes overtime for the matched (using PSM) programme non-participants are assumed to be appropriate counterfactual for programme participants. A decisive advantage of the PSM-DID estimator (conditional DID estimator), compared to a conventional DID estimator, is that this method allows for better control of a selection bias in both observables and unobservables. As shown above, conventional DID methods fail if group of programme participants and a control group are on different development trajectories. Applying a conditional DID estimator (PSM-DID) to measure the effects of a given RDP may greatly improve evaluation findings compared to a situation where a standard PSM (e.g. to estimate ATT) that uses post-intervention data only, is applied.
Advantages	○ In comparison with conventional DID, the combined PSM-DID methods enables a much better control of the selection bias (in observables). ○ In comparison with conventional, binary PSM, the combined PSM-DID method enables a much better control for selection bias (in unobservables). ○ PSM is a very effective tool applied in impact evaluation of various programmes. ○ PSM is a very effective tool to find appropriate control groups (counterfactuals) and estimate impacts/results of individual policy interventions at various levels (individual, farm, food processor, community or region). ○ PSM is a semi-parametric method and a very flexible approach, imposing fewer constraints on the functional form of the treatment model, as well as fewer assumptions about the distribution of the error term. ○ Binary PSM does not require any functional form of assumptions for the relationship linking the outcome variable with the covariates. ○ PSM allows to compute the change in ATT (Average Treatment Effect on Treated), ATE (Average Treatment Effect), ATNT (Average Treatment Effect on Non-Treated) indices straightforwardly.
Disadvantages	○ PSM-DID will not generate reasonable results if other important observable characteristics explaining differences between programme beneficiaries and control group were not included in the model and if the effect of unobservables was not constant over time. ○ PSM-DID requires a relatively large amount of data about supported units and comparison groups before and after the programme. ○ The effectiveness of PSM-DID depends on having a large number of participant and non-participant observations in order to find a substantial region of common support. ○ The outcome of PSM-DID is only true (external validity is large) if the region of common support is large. ○ PSM-DID is a data-hungry procedure: its conclusions only hold on the subset of matched units.

Technique	The combined PSM and DID method
	○ The external validity (generalisability) of PSM-DID results decreases when the share of unmatched units increases.
When to use the technique	The method can be applied to assess programme effects at both micro, regional or programme area levels. However it requires abundant data on the main characteristics and economic performance of both programme beneficiaries and non-beneficiaries, prior and after the programme. An important assumption beyond this method is that the most important characteristics determining programme participation and performance can be observed.
Preconditions	○ Abundant data on programme participants and non-participants prior and after the programme (panel data). ○ High quantitative skills of the evaluator. ○ Good understanding of conditions determining the probability of participation in the programme.

Technique	Generalised PSM method (GPSM)
Description	The technique is used when all units are programme beneficiaries. Given explicit information on the intensity of investment support (e.g. financial flows into a public investment programme per farm, holding, community or region) programme effects (results/impacts) under this framework can be analysed by means of a dose-response function and derivative dose-response function. GPSM method not only allows to estimate the average effect of public investment support on the selected result/impact indicator (e.g. GVA/farm or GVA/region), but also to assess the marginal effects of the programmes or measures, in dependence on the support intensity level obtained. Obviously, such disaggregated programme evaluation results cannot be obtained by employing traditional techniques, e.g. the binary propensity score matching methodology, regression discontinuity design, or any other techniques utilised in standard evaluation studies. It has also been shown that the GPSM method, as an extension of a binary PSM method, can be used to eliminate any bias associated with differences in the covariates included in the evaluation model.
Advantages	○ Very effective tool applied in impact evaluation of various programmes if almost all units receive programme support (at various intensity levels). ○ Very effective tool for finding counterfactuals. ○ It does not require the existence of units which did not receive programme support. ○ GPSM method not only allows to estimate the average effect of public investment support on the selected result/impact indicator but also to assess the marginal effects of those programmes/measures, in dependence on the support intensity level obtained. ○ GPSM is a very flexible approach concerning the functional form of the treatment model (the latter can be empirically tested).
Disadvantages	○ GPSM method will not generate reasonable results if other important observable characteristics explaining differences between various programme beneficiaries were not explicitly included in the model.

Technique	Generalised PSM method (GPSM)
	○ Estimating programme effects using the GPSM method is more difficult compared to PSM or PSM-DID. ○ GPSM requires high analytical skills.
When to use the technique	The GPSM approach is more general than PSM and can be used in situations when all units received support from the given programme and the intensity of this support per unit is known (it also allows to take into consideration units which did not received support, i.e. level of intensity =0). The method can be applied to assess programme effects at both micro, regional or programme area levels. However it requires abundant data on the main characteristics and economic performance of programme beneficiaries (and non-beneficiaries), prior and after the programme. An important assumption beyond this method is that the most important characteristics determining programme participation and performance can be observed.
Preconditions	○ Abundant data on programme participants (and non-participants) prior and after the programme. ○ High quantitative skills (incl. econometrics) of the evaluator. ○ Good understanding of conditions determining the probability of participation in the programme.

Non-experimental design

The instrumental variable (IV) method directly deals with a problem of selection on unobservables. An instrumental variable helps identify the causal effect and must have two characteristics: (i) it must be correlated with programme participation, i.e. it affects the decision to apply to the programme; and (ii) it may not be correlated with outcomes (except through programme participation) or with unobserved variables correlated with the outcome of interest. The choice of the instrument is the most crucial step in the implementation of this method, and should be carefully motivated by economic intuition or theory.

Common sources of instruments may include:

- Policy geographical variation (For instance, if for exogenous reasons, the policy is only implemented in some regions and not in others, so that only part of the population is exposed to the policy).
- Exogenous shocks affecting the timing of policy implementation (For instance, if for exogenous reasons, the implementation of the policy was delayed in one region or for some group of the population).
- Policy eligibility rules (If the policy is designed so that some units are eligible while others are not, parallel with the Regression Discontinuity Design method).

A detailed description of the implementation steps while using IV method to assess programme effects and the list of relevant literature on this subject can be found in (Loi and Rodrigues, 2012).

Main advantages:

- IV method is directly concerned with the selection based on unobservables, either time invariant or not.
- If a proper instrument is found, IV method guarantees that the estimated effect is causal.

Main disadvantages:

- Very difficult to find an instrumental variable satisfying the two crucial assumptions.
- In case of heterogeneous treatment effects, the IV method does not estimate Average Treatment of Treated (ATT), but a local average treatment effect (LATE).
- The instrument may have insufficient variation in the selection of the treatment.
- Estimating programme effects using IV approach requires high analytical skills (econometrics).

When to use the approach:

Considering possible heterogeneity in programme impacts IV methods can be used to account for unobserved selection at the individual level. The IV approach can be applied to assess programme effects when one can find and select good instruments (explicit variables) satisfying the conditions mentioned above. Estimating programme effects requires econometric estimation (2-stage least squares, panel estimates with fixed effects, etc.) of a functional form with programme participation as endogenous variable. The IV approach also involves testing the importance of endogeneity, etc. and requires abundant data (cross-sectional or panel).

Preconditions:

- Available cross-sectional or panel data, on programme participants and non-participants.
- High quantitative skills (incl. econometrics) of the evaluator.
- Good understanding of conditions determining the probability of participation in the programme.

Techniques used in the non experimental design

Technique	Structural models and other modelling approaches
Description	Structural models and other modelling approaches may range from single equation to multi equations, multi country and multi product models. A structural model can be used to estimate unobserved economic or behavioural parameters that could not otherwise be inferred from non-experimental data (e.g. price elasticities, returns to scale, etc.). In programme evaluation, the application of structural modelling approach permits the computation of policy simulations conditional on a set of hypotheses concerned (e.g. preferences and technology). Structural and similar types of modelling approaches (incl. econometric input-output, or CGEs) have been mainly applied to *ex ante* evaluations to determine how different programmes are interlinked with the behaviour of beneficiaries to better understand mechanisms and forecast the potential effects of the programmes under different economic environments. They are only partly applicable to *ex post* evaluation, unless one can draw on already existing models tailored to the rural development context.
Advantages	○ Structural and similar models enable an explicit specification of interrelationships between endogenous variables (such as farm outcomes) and exogenous variables or factors, including budgetary expenditures on programmes. ○ Structural models and other similar approaches provide a possibility to estimate both anticipated as well as non-anticipated programme effects. ○ They also allow to estimate direct effects (at the beneficiary level) and indirect effects (generated from the supply of materials, goods and services attributable to others linked and/or not directly benefiting units and/or industries) as well as induced effects (i.e. multiplier effects) of a given programme generated through direct and indirect activities (including consumption, taxes, etc.) related to a given programme.

Technique	Structural models and other modelling approaches
	○ The above models are usually subject to consistency checks through micro, macro consistency equations.
Disadvantages	○ Input-output models assume that technological or economic relationships are fixed over time and do not respond to price/cost changes. ○ While input-output tables are normally available at relatively high aggregation levels their rescaling to a local level requires using various (often non-transparent) procedures[130]. ○ Commonly applied CGE models usually do not show a detailed enough level of sector disaggregation (a major problem in evaluating RD policies) and are usually static (by contrast, multi-sector and regional dynamic CGE models are much more complex and time consuming in their construction and are therefore very rarely applied to policy evaluations at regional levels). ○ Empirical CGE modelling at regional level often is impossible due to the lack of relevant statistical data and parameters valid at the local or regional level. ○ Estimating structural parameters in these models is usually not straightforward. ○ In CGE modelling heterogeneity of firm behaviour is largely ignored. ○ Modifying CGE or other structural models to reflect local circumstances is usually a considerable and highly time consuming effort that cannot be undertaken by a few external evaluators alone, but requires a great dose of cooperation with the modelling community, local authorities and local stakeholders. ○ The complexity to use the models undoubtedly requires a high level of quantitative expertise and skills. ○ Applying the above models, problems with timeliness of the key data including input-output tables raises questions regarding the results validity.
When to use the technique	Structural and other models are mainly used in *ex ante* assessments of impact/results of a given programme. Under *ex ante* setting, estimating programme effects is carried out through the introduction of model exogenous shocks imitating the programme budgetary outlays. The use of these models to *ex post* estimation of programme macro-economic effects is possible only under certain conditions, e.g.: data used in these models reflects real programme results/impacts at micro-level (the latter could also be derived from other models, e.g. using counterfactual approach at micro-level), assumed transfer mechanisms of external shocks throughout the economy (e.g. various elasticities) reflect a real situation in the economy at the beginning of the programme, etc. Moreover, the evaluator usually faces numerous problems while trying to adjust existing macro-economic models to the needs of evaluation at a regional level.

[130] These can be divided in three main categories: “survey”, “non-survey” and “hybrid” approaches

Technique	Structural models and other modelling approaches
Preconditions	○ Existence of well tailored regional – or macro models to the needs of *ex post* evaluation. ○ Updated information and data reflecting a real situation during the implementation of a given programme. ○ High quantitative skills of the evaluators.

Naïve quantitative approaches

In the past assessing socio-economic impacts of RDP in EU countries was mainly based on quantitative, so-called "naïve" approaches Under this framework comparison groups were selected arbitrarily, leading to quantitative results that were statistically biased (i.e. selection bias). Moreover, in the majority of evaluations, knowledge about a specific programme, direct and indirect effects, e.g. substitution, displacement, multiplier, etc. was "imputed" using non-rigorous methods on the basis of anecdotal evidence or *ad hoc* surveys of a group of beneficiaries, opinions of administrative officials, etc. As we show below, these techniques are in general unsuitable to address appropriately a number of issues generally considered crucial in any quantitative evaluation framework, i.e. construction of relevant control groups, estimation of counterfactual outcomes, or the estimation of the programme general equilibrium effects (e.g. displacement or substitution effects). Below (and for completeness) we present an overview of these methods, and yet believe that an empirical application of those approaches, although straightforward, may lead to producing unrealistic assessments of programme results/impacts.

Techniques used in the naïve quantitative approaches

Technique	Naive "before/after" estimator
Description	Naive "before-after" estimator uses pre-programme data on programme beneficiaries to compute programme outcomes for programme participants (without counterfactual).
Advantages	○ Simple to apply. ○ Relatively inexpensive approach. ○ Requires data on beneficiaries only.
Disadvantages	○ The basic problem with this approach is that information about a change of the outcome of interest (e.g. result indicator, e.g. profits, employment, etc.) in period before and after the programme is arbitrarily attributed to the effect of the RDP. ○ The strong and rather unjustifiable assumptions of this evaluation technique are: - In the absence of a policy intervention (RDP) the result indicator of the programme participants would have been the same as before the programme; - Changes in the outcomes of the programme participants are not affected by any other factor (e.g. macroeconomic, regional etc.) but are the effect of the RDP only.
When to use the technique	This approach "before-after" is NOT recommended to assess an RDP. The approach can only be used if above assumptions are proven (extremely rare situation). Although it is obvious that over the years specific indicators, e.g. gross income or profits do not remain unchanged, some evaluators assign the entire effect of observable change in an indicator of the programme. By doing so the real effects of a given programme may be both

Technique	Naive "before/after" estimator
	under or overstated
Preconditions	○ Data on results observed for programme beneficiaries prior and after programme.

Technique	Naïve "With" vs. "without" approach
Description	Naïve "with" vs. "without" approach is another technique commonly characterised as a naive evaluation approach using all non-participants as a control group. The approach relies on the assumption that in the absence of the programme, the outcome indicator of programme participants would be the same as for programme non-participants. Yet, this would only be justifiable if the systematic performance of programme participants was identical with the outcome performance of programme non-participants. Had this not been the case, the selection bias that results from using outcomes of non-participants as proxy for the outcomes that programme participants would have experienced had they not participated can be very substantial[131]. Obviously programme effects calculated on the basis of this method are unrealistic.
Advantages	○ Simple to apply. ○ Relatively inexpensive approach.
Disadvantages	○ The approach relies on the strong assumption that in the absence of the programme the outcome indicator of programme, participants would be the same as for programme non-participants. ○ No selection bias is assumed.
When to use the technique	The approach "with-without" is NOT recommended to assess an RDP. The approach can only be used if above assumptions are proven (extremely rare situation).
Preconditions	○ Available data on programme beneficiaries and non-beneficiaries.

Technique	Naïve comparison of programme beneficiaries with a population's average
Description	The naïve comparison of programme beneficiaries with a population's average is another simple estimator commonly applied in empirical evaluation studies of RDP. A comparison with a population average uses a control group constructed as a population average (i.e. consisting of programme participants and non-participants). In this evaluation technique necessary data on average outcome indicators in the group of non-participants is usually obtained from various national surveys (or aggregated national data). The approach relies on the similar assumption as when comparisons with all programme non-beneficiaries that in the absence of the programme the outcome indicator of programme participants would be the same as the average of a joint group of programme participants and non-participants. This however would only be justifiable if systematic performance of the group of programme participants (measured by any arbitrary outcome indicator, e.g. income, profit or employment) was identical with the performance of the joint-group of programme participants and non-participants (population average). The effect based on the comparisons of a

[131] See: Heckman, Ichimura, Smith and Todd, 1998

Technique	Naïve comparison of programme beneficiaries with a population's average
	treatment group to an aggregated comparison group of individuals can lead to fairly misleading results[132].
Advantages	○ Simple to apply. ○ Relatively inexpensive approach.
Disadvantages	○ The approach relies on the strong assumption that in the absence of the programme, the outcome indicator of programme participants would be the same as for the joint group of programme participants and non-participants (population average). ○ No selection bias is assumed.
When to use the technique	The approach "comparison with population's average" is only conditionally recommended to assess an RDP. The approach can only be used if the above assumptions are proven (very rare situation).
Preconditions	○ Available data on programme beneficiaries and population average.

4.3.3.3 Qualitative approaches

Qualitative methods should be used in evaluations of EU socio-economic programmes because[133] :

- We are interested in explaining causal patterns. In order to learn from and replicate development, we need to understand what happens inside the black box, to go beyond inputs and outputs. Otherwise we may know what works but not how or why it works. This requires qualitative analysis
- We are interested in impacts for different groups. Programmes often have different impacts for different groups of intended beneficiaries. Breaking down aggregated populations into often quite small groups allows us to investigate these differential impacts.
- We are interested in subtle processes. For example, the effectiveness of networks and partnerships. These are subtle, qualitative phenomena that need to be captured in similarly fine-grained ways.
- We are interested in contexts. These are made up of many different factors, geography, culture, economic structures, social groups, institutional arrangements, climate, employment patterns, past development, etc., and the way they interact in particular development settings can only be described in qualitative terms. Furthermore the entire development process needs to be set into context if lessons are to be learned that will be transferable.
- We are interested in human judgements. These may be the judgements of stakeholders whose intervention logics and programme theories evaluators want to elicit. Or they may be the judgements and experiences of the intended beneficiaries of socio-economic development.
- We are interested in bottom-up understandings. These can include: the development ambitions of grass-roots actors (small firms, municipal authorities, professional associations) and the expectations and experiences of local people in a local development setting.

Qualitative analysis can be extremely useful in providing a preliminary assessment of potential effects expected to be generated by a given RDP. They are also invaluable in generating information which may be critical to understand mechanisms leading to these effects. The latter may include qualitative information and understanding of local circumstances under which programme was implemented, e.g.

[132] Lechner (2001)
[133] See: Evalsed: The resource for the evaluation of socio-economic development, EC, September 2013 available under:

local socio-cultural and economic context or specificity of institutional settings (formal and informal), but also important details on the specificity of the group of beneficiaries which can be later used in quantitative assessments. The evaluator can apply qualitative methods to provide context and explanations for quantitative results, to explore "outlier" cases of success and failure, and to develop systematic explanations of the programme's performance as it was found in the quantitative results[134]. Qualitative methods can also be used for programme evaluations concerned with provision of public goods.

The approach taken in qualitative research applied for programme evaluation is markedly different than in quantitative research. In general, the most important characteristics of a qualitative research applied in programme evaluations are[135]: Focuses on meaning, and on 'the why'; Tends to focus on new issues where understanding is required, rather than on confirming prior hypotheses; Seeks to discover the 'insider' rather than 'outsider' perspective; Is person-centred rather than variable-centred; often taking a humanistic or phenomenological perspective; Often seeks to promote joint learning by all participants, rather than just the learning of the researcher; Is holistic rather than particularistic; Is contextual (situated, embedded) rather than de-contextual (distant, removed or detached)and; Seeks to understand depth rather than breadth.

Main advantages:

- Highly applicable for the analysis of a *process* of change;
- Flexible and can be specifically tailored to the needs of the evaluation using open-ended approaches, can be carried out quickly using rapid techniques, and can greatly enhance the findings of an evaluation through providing a better understanding of stakeholders' perceptions and priorities and the conditions and processes that may have affected program impact[136];
- Can explain turning points and changes over time in causal patterns established with quantitative data;
- Invaluable in generating information which may be critical to understand mechanisms leading to programme effects.

Main disadvantages:

- Subjective and therefore characterised by personal biases;
- Highly dependent on the skills of the investigator who makes principal decision on what to ask and what inferences to make;
- Requires significant time and interpersonal communication expertise in setting up interviews and focus groups;
- Characterised by limited external validity, i.e. outcomes lack statistically robustness and are not easily transferable from case studies to an overall population;
- Difficult to credibly measure outcomes by establishing a counterfactual (more difficult to indicate what would have happened in the absence of the RDP/measure) and in consequence more difficult to establish reliable programme effects;
- Qualitative interviews are generally not conducted randomly. As a result, the qualitative interviews often involve persons who are more likely to participate in the programme (e.g. direct programme beneficiaries), thereby leading to a bias in assessing the real programme effects, results and/or impacts;

[134]See: World Bank, 2011
[135]See: Vanclay, 2012
[136]See: Baker, World Bank, 2000

- Conclusions based on qualitative methods can be especially complex and even contradictory.

When to use the approach:

Qualitative evaluation techniques are particularly useful during the three stages of an evaluation[137]:

- When designing an evaluation, the evaluator can use focus groups and interviews with key informants to develop hypotheses as to how and why the programme worked and to clarify research questions that need to be addressed in the quantitative evaluation work.
- In the intermediate stage, before quantitative evaluation results become available, qualitative work can help provide policy makers quick insights into what is happening in the program.
- In the analysis stage, the evaluator can apply qualitative methods to provide context and explanations for the quantitative results, to explore "outlier" cases of success and failure, and to develop systematic explanations of the programme performance as it was found in the quantitative results[138].

In that sense, qualitative work can help explain why certain results are observed in the quantitative analysis, and it can be used to get inside the "black box" of what happened in the programme[139].

a) Introductory qualitative analysis to capture basic trends in programme effects

How the programme impact is defined will necessarily determine the scope and context of *ex post* evaluation. Given that impacts of RDPs are defined broadly, i.e. "positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended", it is advisable that at the initial stage of *ex post* programme evaluation the evaluator takes advantage of an introductory qualitative analysis, e.g. in the form of systematic structured interviews with programme stakeholders to capture basic trends in "experienced" programme effects. The objective of the introductory qualitative analysis should be to prioritise different aspects of potential programme impacts by indicating and selecting those which, in view of practitioners, beneficiaries, partners and policy makers, have been making a biggest "change" in comparison with a situation "without" the programme. By suggesting areas where programme effects are multi-dimensional, i.e. including unintended or uncertain effects, the evaluator and the stakeholders can draw on an earlier developed theory of change. At this stage, an indicative analysis can also include a preliminary assessment of the scope of "observed" effects which at a later stage should be a subject to more rigorous quantitative verification using causal inference approaches. Clearly, at the first stage of evaluation, preliminary qualitative assessments are essential because they can provide invaluable insiders' perspectives on a programme performance (especially regarding its unintended results or various side effects).

b) Complementary explanations on programme performance

In evaluation, qualitative research is used specifically to consider the why and how questions that quantitative methods typically cannot answer[140], e.g.:

- Why does the programme work (or not work)?
- How does the programme achieve its goals?

[137] See: World Bank, 2011
[138] Although the applicability of qualitative methodologies to construct valid counterfactuals is considered as rather limited. Nevertheless, recently developed more rigorous qualitative methodologies provide some encouraging examples for approaching this problem, see: Cummings, 2006; Hind, 2010.
[139] See: Bamberger, Rao, and Woolcock, 2010
[140] A blurring of the boundary between quantitative and qualitative methods follows when we distinguish between methods to gather data and methods to analyse them. Data gathered can be qualitative, e.g., interviews, surveys, focus groups and still be analysed quantitatively. Many statistical models for example use interview or questionnaire data as inputs. And quantitative analyses may only be fully understood after qualitative interpretations of what the results mean. – see: EC, DG REGIO, 2013

- Why does it work for some clients (or in some situations) and not for others?
- What are/were the needs of the clients that were not anticipated by programme developers?
- What were the additional unintended and/or unexpected positive or negative consequences of a given programme?

Enabling counterfactual analysis in case of missing data

Furthermore, when there is a small number of observations (a created comparison group is not able to be statistically founded), it is important to use some of the systematic and rigorous non-experimental qualitative methods developed in fields like political science (e.g. Qualitative Comparative Analysis) enabling address the challenge of casual inference (see: International Initiative for Impact Evaluation, 2014).

In the past, qualitative techniques were mainly used for carrying out evaluation with the intent to determine impact by the reliance on other approaches than the counterfactual to make a causal inference (Mohr 1995)[141]. The focus was on understanding processes, behaviours, and conditions as they are perceived by the individuals or groups being studied[142]. Because measuring the counterfactual is at the core of impact analysis techniques, qualitative designs have generally been used in conjunction with other evaluation techniques[143].

More recently, there is an increasing tendency to apply qualitative methods also to a quantification[144] of various results/impacts of a given programme, especially if: i) programme spill-over effects are widely diffused, ii) control groups not affected by a given programme cannot be easily found, and iii) application of specific quantitative methodologies is not considered as feasible. In such a case qualitative approaches may be helpful. Yet, they should always be applied in a rigorous manner, e.g. in the frame of a programme theory-based evaluation.

Preconditions:

- Knowledge of multifaceted techniques applied in qualitative approaches.

Techniques used in qualitative methods

Technique	Focus groups
Description	Focus groups consist of a small number of individuals brought together to discuss a topic of interest. Focus groups should be facilitated by an external moderator and comprise not less than 7 and not more than 15 people, a far as possible stemming from different sub-groups of stakeholders (e.g. Managing Authority, implementing body, beneficiaries, independent experts, etc.). A focus group usually comes together several times; it can as well be installed as a core element for the continuance of an ongoing evaluation.
Advantages	○ Focus groups allow to verify collected qualitative and quantitative evidence with various types of stakeholders at the same time. ○ Thoroughly constellated and well facilitated focus groups may reflect the universe of programme stakeholders in a nutshell. The focus group added value is not only provided by the reflective capacity of the individual members, but by the group intelligence (group intuition), bringing forth new hypotheses and insights on critical factors and interrelationships, providing valuable lessons about patterns of success

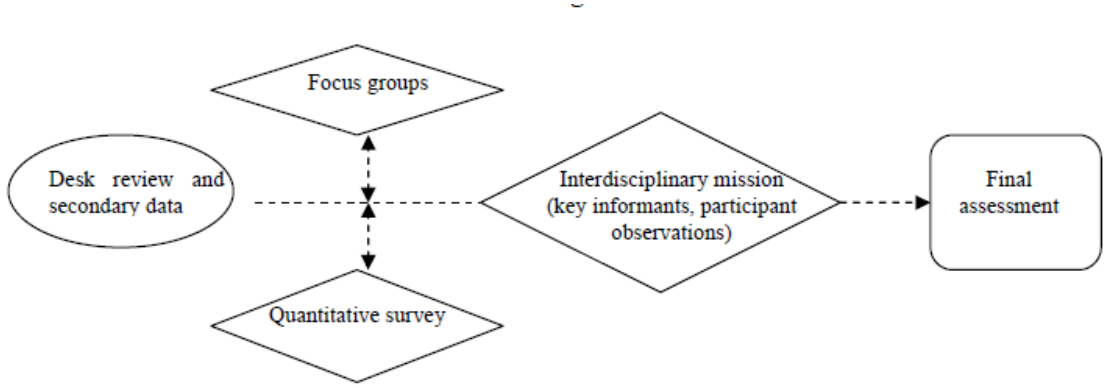
[141] Baker, World Bank, 2000
[142] Valadez and Bamberger 1994
[143] World bank, 2000
[144] This is done, for example using an arbitrary scale, e.g. from -5 to +5; or by comparing a situation in a given locality to a certain benchmark (e.g. EU average), etc.

Technique	Focus groups
	and failure. ○ Focus groups are indispensable instruments for institutional learning (at the level of programme design and policy shaping). As such focus groups are an intervention as such, by strengthening (self-) reflective capacity of a local partnership, specifically in the context of Leader. As for the assessment of impacts or results, focus groups produce a host of practical insights and help to sharpen the evaluation judgements; for example, the value added created by a direct marketing activity could be juxtaposed to the increase of workload on the female farming people, thus making a link between the gross value added and the quality of life indicator. ○ Finally, the area-based perspective of focus groups implies that it could discover important realms which are not observed by the existing evaluation device. In this event it could propose and establish programme-specific indicators.
Disadvantages	○ Focus groups involve only a small number of people and topics. It is therefore of utmost importance to pick the "right people", to find the right mix, and to avoid "one-way thinking" in the selection of participants. ○ Focus groups only "sketch" the topic at a certain moment of time. Unless they are repeated across time, they cannot provide a picture on the trends and tendencies.
When to use the technique	Focus groups should be used whenever it is important to verify collected quantitative and qualitative evidence with various types of stakeholders and thus triangulate evaluation findings.
Preconditions	○ Identification of the most relevant participants. ○ Carefully prepared method to run the focus group. ○ Design relevant key questions to answer.

Technique	Surveys
Description	Surveys are mixed methods (qualitative and quantitative) applying a deductive analytical approach. Surveys deduct the information from a (representative) sample in order to depict reality of the total. In a deductive approach, hypotheses are mostly built from literature before surveys are designed. In a survey the information is gathered from (a sample of) entities for the purpose of constructing quantitative descriptors of the attributes of the larger population of which the entities are members.
Advantages	○ Allows the collection of primary data. ○ Bridges existing data and information gaps. ○ Can collect both qualitative and quantitative data.
Disadvantages	○ Represents an *ad hoc* tool to collect data and information. ○ Data and information are collected from selected samples of beneficiary/non-beneficiary population. ○ Requires a lot of resources.
When to use the	Surveys are to be applied in various circumstances of the assessment of effects,

Technique	Surveys
technique	dependently on the type of information to be collected (e.g. sampling of control groups), the type of indicator it is serving (social science adopts different survey methods than natural science), the programme specifics (data availability, type of RDP measure). Surveys can also be very valuable in collecting quantitative data (e.g. basic characteristics and performance of supported and non-supported units) which can be used later for an estimation of programme results based on e.g. matching techniques. While using this data collection technique a particular attention has to be paid to the accuracy of information concerning the past events, e.g. characteristic and performance of programme supported and non-supported units in a time period prior to the implementation of a given RDP/measure (a double check may be required). There is a vast amount of literature in this field which allows for further reading and filtering of the appropriate methodology. However, there is no "one-method-fits-all" survey methodology to be identified. Thus the application of surveys related to the different thematic fields of indicators has to be deliberately selected and accordingly adjusted. A representative survey of programme components can be used together with questionnaire-based interviews of programme beneficiaries and non-beneficiaries; and in-depth interviews and focus groups interviews to gather preliminary evidence of programme effects. Under this approach focus group discussions may help to identify possible sources of bias in a quantitative survey and a way to address them. A schematic flow chart of such an approach is shown in Figure 11 below.
Preconditions	○ Design the appropriate questionnaire with relevant questions and their proper combination. ○ Identify and reaching the relevant target groups. ○ Design the system of collection and elaboration of survey results.

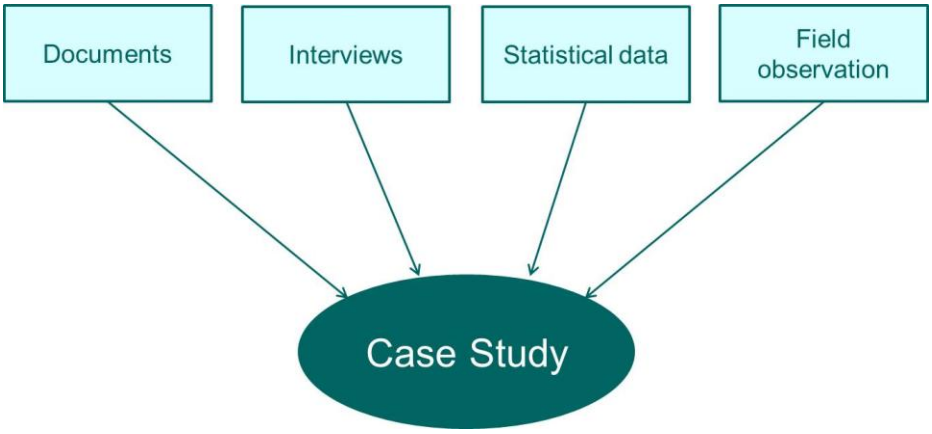
Figure 11 Example of using surveys in combination with various qualitative and quantitative methods in the assessment



Source: NONIE evaluation guidelines, 2008

Technique	Case studies
Description	Case studies are the preferred evaluation tool when "how" and "why" questions are being posed. Case studies allow a detailed examination of the actual elements in line with the evaluation goals. The purpose of the case study is to provide a picture, which is often more reliable than the outputs from other tools in the context of the scarcity of basic data (which is often the case in country evaluations).The Success Case Method identifies individual cases that have been particularly successful (and unsuccessful) and uses case study analytical methods to develop credible arguments about the contribution of the intervention to these. Normally a standard case study includes the following elements: i) Examining and discussing intended and non-intended socio-economic effects; ii) gathering evidence of these effects; iii) assessing causal contribution; and iv) drawing conclusion on managing evaluation.
Advantages	○ Provides in depth illustration of the situation in the particular context when being the subject of evaluation. ○ Triangulates evaluation findings based on qualitative or quantitative methods. ○ Helps to better understand the causality observed with the "dry" numbers.
Disadvantages	○ Only allows for a limited number of observations.
When to use the technique	Case studies can be used in qualitative as well as quantitative assessment, drawing from several established traditions including theory based evaluation, organisational development, appreciative inquiry, narrative analysis and quantitative statistical analysis of effects[145]. Case studies serve to underpin or discard hypotheses, and due to their idiographic character they are rather useful to illustrate or to exemplify results which already seem to be sufficiently confirmed. With their help, intrinsic driving factors and cause/effect relationships can be made explicit and thus become accessible for discussion. In this respect case studies could be used in the assessment of various programme effects, e.g. the cross-effects of axis 2 measures on farm employment (axis 1) or the quality of life in rural areas (axis 3). They could also help to analyse which type of non-farming but farm-based employment (e.g. IT-based activities, tourism or social services) interrelates in which way with farming activities.
Preconditions	○ Good knowledge of the programme beneficiaries, their geographical distribution and economic situation. ○ Existence of good examples to be elaborated as representative cases. ○ Good case study writing technique.

Figure 12 Main components of case study



Source: based on EuropeAid, EU

Other qualitative approaches can be used especially for gauging evidence in the field of capturing aspects of "quality of life" or "attractiveness of an area" (most relevant for Axes 3 and 4), when additional methods are useful. One of these approaches may be the Critical Incident Technique: In this approach individuals describe events relating to the impact of a specific RD policy intervention they found to be especially satisfying or dissatisfying. The first step determines and reviews the incident, then fact-finding, involving the collection of details of the incident from the participants. When all the facts are collected, the next step is to identify the issues. Afterwards a decision can be made on how to resolve the issues based on various possible solutions. The final and most important aspect is the evaluation that will determine if the solution selected will solve the root cause of the situation and will cause no further problems.

Another approach for capturing subjective impressions of effects of RD measures would bet o provide tools for "looking through the eyes of the beneficiaries". There are various methods for achieving this:

- Film assessment – the effects of the RDP are assessed through asking beneficiaries to produce a film depicting the most important achievements of the programme for them in their programming areas. A control group may be established by asking the same to non-beneficiaries.
- Photo-galleries – beneficiaries may be asked to provide a photo gallery of the most positive/negative effects of RDP.

These types of methods are especially useful to answer the "ultimate" evaluation question within RD evaluations – i.e. whether RDP have met the "needs" of the population within the programming areas, as "population needs" are subjective and cognitive constructs not to be captured by standard indicators alone.

4.3.3.4 Mixed-methods approach

The mixed-methods approach[146] is highly applicable to evaluation of the RDP and can be described as an application of a mix of quantitative and qualitative data sources, types of data, sampling methods, analysis methods in order to balance the limitations of any method. Mixed methods evaluations seek to integrate social science disciplines with predominantly quantitative and predominantly qualitative approaches to theory, data collection, data analysis and interpretation. Mixed-methods approaches develop more comprehensive evidence of programme results and impacts and for example can be

[146] For a guidance and overview of applications see: Bamberger M. "Introduction to mixed-methods in impact evaluation", Impact Evaluation Notes No 3, August 2012; Bamberger, Michael, Vijayendra Rao, and Michael Woolcock. 2010. "Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development." Policy Research Working Paper 5245. World Bank, Washington, DC.

used to distinguish between implementation failure and theory failure in the event of a lack of intended results/impacts[147]. The key strength of this approach is the ability to provide triangulation of methods and data sources (e.g. by addressing a weakness of one method by using another technique). One of the main areas in the RDP *ex post* evaluation where mixed-methods are especially applicable is verification and an in-depth analysis of the main reasons for high, medium or low effectiveness of programme support. If the results from different methods converge, then inferences about the character and magnitude of these impacts will be stronger. If they diverge, they can provide a more objective explanation of factors behind it.

Main advantages:

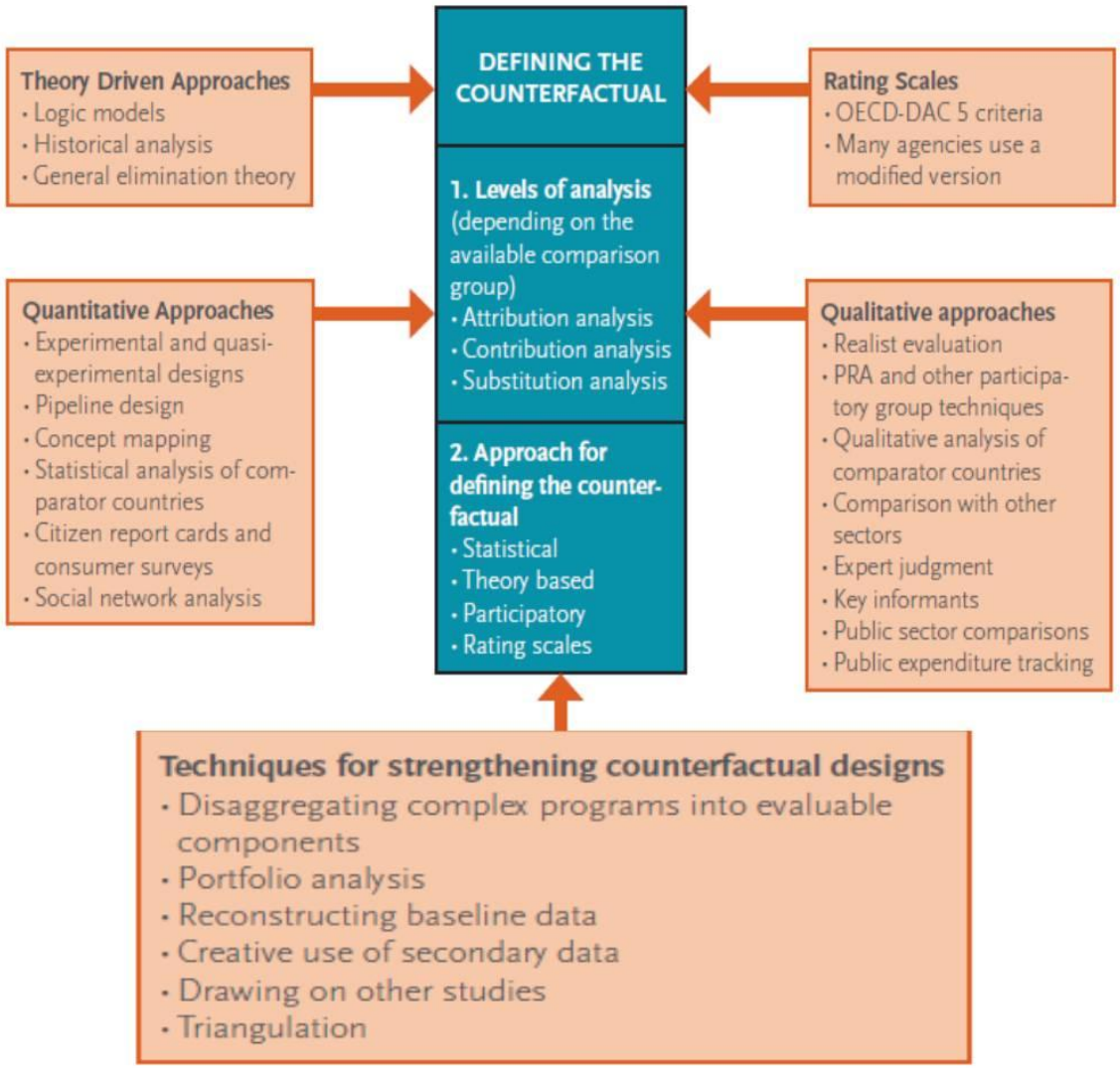
- Strengthen the quality of evidence by compensating for the limitations of any source through a complementary mix of research questions, process for developing research questions, sampling procedures, data collection procedures, type of data, type of data analysis.
- Provide a triangulation of methods and data sources, e.g. by addressing a weakness of one method by using another technique.
- Support respect for different ways of thinking and knowing.
- Extend the comprehensiveness of evaluation findings through results from different methods that broaden and deepen the understanding reached.
- Provide a unique opportunity for an intentional inclusion of diverse (stakeholder) perspectives and an appreciation for context and complexity.
- Incorporate a wider diversity of values through the use of different methods that themselves advance difference values. This encourages greater consciousness about the value dimensions of the evaluation.
- Can improve the rigour of causal analysis by combining and cross-checking performance of various techniques.
- Is especially useful in gaining a comprehensive view of the programme's effectiveness.

Main disadvantages:

- If not implemented effectively, can lead to a lack of coherence or increase dissent between different perspectives.
- If not correctly and rigorously applied, can provide a misleading assessment of credible evidence.

[147] NONIE, 2008

Figure 13 Using mixed methods design for evaluating complex interventions



Source: Bamberger M. (2012),

A mixed-methods approach, i.e. involving the integration of rigorous quantitative and qualitative methodologies is the best empirical approach to assess an RDP. It is recommended that the mixed approach should be carried out during each step of the evaluation[148]. For example, the mixed approach should be used to validate the most reliable counterfactual. Supplementary qualitative analysis can be carried out along with quantitative analysis to explain the main reasons for high, medium or low effectiveness of the programme support in all regions or areas under consideration. While effectiveness can be verified in the form of a summary assessment, what has been planned in comparison to what has been achieved at the level of inputs, outputs, results and impacts, this assessment should be subject to verification using a mixed-methods approach.

A mixed-methods approach is also very effective in ensuring a comprehensive estimation of programme externalities. The major steps might be as follows:

Step 1: Initially assess various expected or unexpected, positive or negative effects of a RDP to start with the theory of change, link it with the programme intervention logic, and apply *an introductory qualitative analysis* to address the scope of "observed" gross effects which at a later stage should be subject to more rigorous quantitative verification using causal inference approaches. The main aim of

[148] See: Baker, World Bank, 2000.

the introductory qualitative analysis should be to prioritise different aspects of potential programme impacts by indicating and selecting those which, in view of the RD stakeholders and policy makers, have been making a biggest "change" in comparison with a situation "without" the programme.

Step 2: Formulate testable hypotheses regarding potential programme effects. This activity should be undertaken at the beginning of *ex post* evaluation. At this stage CEQ and judgement criteria linked to the assessment of the expected outcomes should be complemented with programme-specific evaluation questions, judgement criteria and testable hypotheses derived from Step 1.

Step 3: In this step, above hypotheses should be empirically tested and/or statistically verified using rigorous methodological approaches based on causal inference (both quantitative as well qualitative methods can be employed). At this stage it is important to use meaningful result indicators to enable the analysis of the most important expected or unexpected, positive or negative aspects of programme effects grouped under economic, social and environmental domains. The same indicators should be later employed to test and verify the above hypotheses empirically, i.e. using collected data and applying suitable methodologies (i.e. under rigorous evaluation of effects using credible counterfactuals).

The main components of Step 3 are as follows:

- Defining the outcome variables
- The analysis can be conducted with respect to as many outcome variables there are data for. The analysis can be extended not only to cover direct effects, but also intermediate outcomes, indirect programme effects, etc.
- Defining the time dimension. For example, 2006 or an average 2005-2006 (i.e. reflecting situation prior to the current programme) with 2013-2014 (i.e. situation after the end of the current programme).
- Applying suitable methodologies for finding a credible control group (in principle, both qualitative as well as quantitative methodologies can be here applied).
- Computing an average outcome for the group of programme beneficiaries.
- Computing an average outcome for a comparable control group.
- Calculating the effect of the programme (for each outcome variable separately) using respective Average Treatment Indicators (see below). Average Treatment Indicators can be applied to assess both programme direct as well as specific indirect effects.

The specific techniques applied in mixed-methods approach are (e.g. surveys, case studies) cost-benefit and cost-effectiveness analysis (See **Part II:** Chapter 4.3.3.3 Qualitative approaches).

Cost-benefit analysis attempts to measure the economic efficiency of programme costs versus programme benefits. This method can also be applied in situations where it is not possible to measure all programme benefits in monetary terms. The main steps of cost-benefit and cost-effectiveness analysis are to identify all programme costs and benefits and then compute a cost-to-effectiveness ratio. Benefits can be monetary, yet if benefits cannot be quantified it is possible to use subjective or qualitative indicators such as ranking or weighting systems. This approach, however, can be tricky in interpreting subjective score. According to the World Bank (2000), incorporating cost-benefit or cost-effectiveness analysis to evaluation of programmes and policies is strongly recommended. This methodology can enable policy-makers to compare alternative interventions on the basis of the cost of producing a given output.

Preconditions:

- Knowledge of strengths and weakness of quantitative and qualitative evaluation methodologies.
- Ability to apply rigorous quantitative and qualitative evaluation techniques.
- Good access to quantitative and qualitative data.
- High level analytical capacities of the evaluator.

4.3.4 How to select an appropriate evaluation method

Depending on data availability, budget and time constraints, as well as country circumstances, evaluations of RDPs will be different and will require some combination of appropriate methodologies, both quantitative and qualitative.

4.3.4.1 Meeting basic evaluation standards

The framework for testing different evaluation methods regarding their appropriateness for the assessment of the effectiveness, efficiency and results and impacts of different types of measures under RDP should include the quality criteria that operationalise commonly accepted evaluation standards, for example those described by NONIE 2008.

Assessing the applicability of a given evaluation method is the key step in choosing an appropriate evaluation approach and an important decision that may influence evaluation results to a great extent. The expectation is that any valid evaluation methodology should bring about similar results and reflect the true effect of the programme support. Therefore each chosen method should be subject to critical screening by taking into consideration the basic evaluation standard criteria (credibility, rigour, reliability, robustness, validity, transparency and practicability). Furthermore, selecting a given method should be based on the ability to explain causality, eliminate a possible selection bias, isolate the effect of the programme from other factors, etc. There is no single 'best' method[149] for assessing the effects of policy interventions, yet, some methods have a comparative advantage over others in analysing particular objectives, e.g. quantitative methods (experimental or quasi-experimental methods) have a comparative advantage over other methodologies for large interventions and in addressing the issue of attribution of the effect of a given programme.

While making a selection of an appropriate evaluation method one also has to remember that the majority of evaluation techniques are NOT substitutable with each other. For example, a counterfactual approach that can be applied to the evaluation of programme results at micro-level cannot be replaced by another quantitative input-output method which is only applicable to assess some programme effects at regional or macro-level. Also a counterfactual evaluation methodology (e.g. propensity score matching) applicable to the evaluation of economic results or impacts of a given RDP cannot be replaced by e.g. an environmental impact assessment method, which can only be applied to estimate environmental effects, etc.

In consequence, it is advisable that the evaluator aims at choosing an evaluation design based not on a single method but a bundle of methodological approaches which are internally consistent and supplementary to each other. Moreover, they meet all the basic evaluation quality criteria referred to above.

The evaluator should try to select as strong evaluation design as possible (the latter may consist of a bundle of evaluation techniques and methods), bearing in mind time, money and practicability constraints (including data availability).

[149] NONIE, 2008

Table 8 Example of the assessment of matching methodology using specific quality criteria

Judgement criteria	Assessment	Score (min 1 – max 10)
Ability to explain causality	The matching method can be applied to causality analysis. After the "ideal –Golden Standard" i.e. experimental design, PSM and GPSM are the second best rigorous, credible, reliable, robust and transparent approaches for building counterfactuals.	**9**
Ability to eliminate a possible selection bias	PSM and GPSM are the second best rigorous, credible, reliable, robust and transparent approaches to eliminate selection bias after "ideal" i.e. experimental design.	**9**
Ability to isolate the effect of the programme from other factors	PSM and GPSM are very effective approaches used to isolate the effect of RDP from other factors. Yet, these methods can be less useful in situations where observed result/impact indicators are simultaneously influenced by unobservables which impact is not uniformly distributed between beneficiaries and control groups (e.g. other simultaneously implemented structural programmes may also target beneficiaries of the RDP but do NOT equally affect performance of RD control groups).	**8**
Taking into account potential spill-over effects	The matching method can be useful only to assess certain spill-over effects at micro-level (e.g. leverage effects) or substitution effects at regional level. On the other hand the GPSM methodology (i.e. applied directly at regional or macro-level) can be helpful in assessing various types of spill-over effects (in rural areas which were directly and not directly targeted by the RDP.	**6**
Validity check	The matching method is one of the most credible, reliable, robust and transparent approaches, yet subject to abundant data and proficient skills and capacities of the evaluator. External validity check: The matching method can to a large extent be utilised to infer about programme effects on the whole population on programme beneficiaries who received support from a given measure. Yet, the external validity of results based on this method decreases when the share of unmatched units is large.	**8** **7**
Validation judgement criteria	The basic justification of applying evaluation methods based on matching is that this design is the closest to the experimental one (ideal benchmark). As a consequence, all other methods will have to be compared to this one. This refers to a) similarity of policy conclusions, and b) requires statistical analysis of differences between different methodological	**9**

Judgement criteria	Assessment	Score (min 1 – max 10)
	approaches.	
Verification 1: applicability of measurement	Methodologies based on matching can be applied to analysis of wide ranges of possible results/impacts using various alternative result/impact indicators (including environmental ones). Methodologies based on matching can also be applied for verification of alternative explanations regarding reaching of obtained results/impacts	**9**
Verification 2: applicability of a given evaluation method in the event of non-existence of data for non-supported farms/regions	No, for binary PSM method. GPSM method allows to estimate results/impacts of a given programme/measure in a situation where all units farms, holdings or regions received support from the RDP(at a different intensity level). The above methodology is rigorous, credible, reliable, robust and transparent (yet, a sensitivity of obtained results could be an issue).	**8**

4.3.4.2 Choosing right evaluation approach given feasibility (time, budget and data) constraints

Conducting rigorous impact evaluations is crucial for policy learning (it is indispensable in providing evidence about what works, what does not and the reasons why?). Furthermore, it is important that evaluation methods, findings and recommendations are as reliable as possible. To achieve this, it is essential to prepare conditions for conducting a sound *ex post* evaluation a long time ahead, the best at the beginning of programme implementation or at least at early stages of ongoing evaluation. Early thinking on evaluation design can facilitate better budget planning, save resources and improve data quality and quantity. If this is not case, the evaluator may face severe budget, time and data constraints which may act as disincentives to conduct rigorous evaluations.

In these circumstances a particular challenge is to select the strongest possible evaluation design enabling capturing advantages of rigorous impact methodologies given above real-world evaluation constraints.

The areas which are the most likely affected by the three evaluation constraints are as follows:

- **Budget constraints** - affect the ability to combine quantitative and qualitative data collection and analysis, the size and professional experience of the research team, and the analysis that can be conducted.
- **Time constraints –** affect, when the evaluation begins and ends, how long researchers can be in the field, and time available for feedback from the stakeholders.
- **Data constraints** affect the ability to collect information from a suitable comparison group and collection of sensitive information and/or interview groups that are difficult to reach groups. When evaluation is based on secondary data or when data is obtained through studies conducted by other agencies, it may affect the compatibility of sample coverage and timing, or whether data covers the required variables and define them in the required manner.

As data collection may represent more than half the cost of an evaluation[150] it may be advantageous to reduce the costs of data collection by simplifying the evaluation design. However, by doing this, one has to be aware that there are two common evaluation designs which do not qualify as sound impact evaluation designs (see: World Bank, 2006, 2010). These are: a) pre- and post intervention comparison of programme beneficiaries, and b) post intervention analysis of programme beneficiaries without a control group.

Table 9 Two frequently used evaluation designs where the lack of an acceptable counterfactual almost always disqualifies them as quality impact designs

Impact Evaluation Design	% costs saving compared with the most rigorous evaluation design	Remarks
Pre- and post-intervention comparison of programme beneficiaries with no comparison group	50	A common design when data is only collected on the project beneficiaries. Methodologically weak as using pre intervention period as counterfactual requires heroic assumptions about time varying effects and individual unobservable variables.
Post intervention analysis of programme beneficiaries without a control group	75-90	The weakest design but one which is commonly used when evaluations have to be conducted late in the programme evaluation with very limited time and money. Qualitative methods, programme records and aggregate secondary data are used to estimate counterfactual.

Source: adapted from World Bank and IEG Group, 2006

Given the above it is therefore important that by facing real world constraints, the evaluator does not retreat to the weakest evaluation designs which are not in a position to generate reliable and robust results (see below).

In general, there are several options for carrying out sound evaluation under budget, time and data constraints. The most important are[151]:

a. Select comparison groups on the basis of easily observable and identifiable characteristics. By selecting this option it is important to keep in mind the potential for unobservable differences, to address them as far as possible through qualitative research, and to attach appropriate caveats to the results.

b. Using as much as possible secondary data. Yet, when working with secondary data one has to take into consideration: i) Time differences between the start of the project (when baseline data are required) and the time when secondary data were collected or reported, ii) How closely does the sample approximate the target population?; iii) Was information collected on all key programme relevant variables and outcome indicators and is data adequate for the purposes of

[150] See: World Bank, 2000
[151] See: Conducting quality impact evaluations under budget, time and data constraints, World Bank and IEG, 2006

the evaluation using as much as possible administrative records collected under the programme.

c. Apply triangulation methods to check secondary data sources.

d. Reduce the number of interviews conducted (accepting a lower level of statistical precision, e.g. 90% instead of 95% of confidence intervals.

e. Reduce. the level of statistical disaggregation in the analysis (e.g. only obtaining results on the whole population instead of computing effects on different groups).

f. Replace individual interviews with community level data collection.

g. Allow sufficient time to meet with key stakeholders.

h. Develop a programme theory model and articulate the effects chain through which impacts are expected to be achieved.

i. When operating with smaller than ideal sample sizes, it is sometimes possible to develop cost-effective mixed-method approaches that strengthen validity by providing two or more independent estimates of key output, result or impact indicators.

j. When working with small samples, mixed-method approaches can provide a cost-effective way to strengthen theoretical framework and the validity of the counterfactual.

k. Use contextual analysis to understand how local factors might affect outcomes.

l. Multivariate analysis should be used where possible to strengthen the match between the programme beneficiary and comparison groups (and to strengthen validity of projections).

In general, if an acceptable impact evaluation cannot be conducted within the above constraints, the resources and time frame must be renegotiated[152].

In former evaluations of RDPs, the issue of data availability - or rather data non-availability - was often regarded as the main constraint deciding upon which evaluation design was selected. Indeed, multiple examples show that the evaluator had to sacrifice methodological rigour, credibility, robustness and validity in data situations which could be described as not optimal. Obviously, the link made in various empirical studies between data availability and the selection of the appropriate method can be regarded as being too strong. The main problem with such an approach (a strong dependence of the selected methodology on data) is that a decision of not using certain rigorous methodologies (for example those which rely on the use of appropriate counterfactuals) and instead applying simple or "naïve" evaluation techniques, "justified" by data "non-availability", could result in introducing a significant evaluation bias to the obtained results – thus making the latter less reliable, not-robust and unrealistic. The relevant question is therefore: what are possible trade-offs between different methodologies in situation when data availability is not optimal?

In order to answer this question one has to look closer at the specificity of each individual method in relation to certain quality criteria, e.g. credibility, rigour, reliability, etc. which determine the overall quality of evaluation on one side, and respective data demand on the other side.

An illustrative example of possible trade-offs between data demand and potential bias in results derived from application of individual methodologies to the assessment of net programme effects is presented in Table 10.

[152] The Real World Evaluation (see: Condensed Summary of Real World Evaluation, 2nd edition, SAGE publications, December, 2011) has recently developed sets of evaluation guidelines for addressing budget, time, data, political, organisational, and administrative constraints and challenges. In order to prevent evaluators from producing evaluations that result in finding and recommendations that are based on "we did the best we could under the circumstances", but are methodologically questionable, defining "Minimum Acceptable Quality Standards for Conducting Evaluations Under Constraints" is under preparation.

Table 10 Comparison of various methodological approaches

Method	Credibility/Rigour/ Reliability, etc.	Ability to reduce selection- and other biases	Quality of evaluation	Data demand
Experimental approach	+++++	+++++	+++++	+++
Quasi-experimental approaches	++++	++++	++++	++++
Matching approaches (combined with DID)	++++	++++	++++	++++
RDD	+++	+++	+++	+++
IV	++++	+++	+++	+++
DID	++	++	++	++
Comparisons with non-beneficiaries in a given period of time (naïve approach)	+	+	+	++
Before-after comparison of programme beneficiaries (naïve approach)	+	+	+	+
Qualitative approaches applied to estimation of programme results/impacts	++	++	++	+

Where:

+++++ = the highest score

\+ = the lowest score

The above table, which is purely illustrative, shows that:

- Evaluation methodologies usually ensuring higher evaluation quality are usually also more data demanding.
- Trade-offs exist between the use of less data demanding evaluation methodologies on one side, and the occurrence of a potential bias in obtained results on the other side.
- This trade-off is especially large in case of naïve and qualitative approaches, i.e. less data demanding methodologies are usually not able to reduce selection and other types of biases and systematic errors, and thus results obtained from these methodologies about programme results/impacts are normally less credible, less robust, less reliable and therefore of a limited use for policy-makers.
- Decision to use or not to use certain evaluation designs (especially those which are based on counterfactuals) and apply less rigorous or simplistic techniques, only because data availability

may be restricted, can be highly distortionary and may lead to producing unrealistic assessments of programme results/impacts.

- An appropriate bundle of methods can minimise this potential bias and systematic errors. In such a situation additional surveys to collect additional necessary data on characteristics and performance of programme beneficiaries and non-beneficiaries instead of taking the risk of overstretching the interpretation of programme results based on simple yet biased evaluation techniques is highly recommended.

4.3.4.3 Suggested methods in relation to Axis 1

In relation to programme-level EQs

Evaluation Question	Recommended approach
1. To what extent has the RDP contributed to the growth of the whole rural economy? (Impact indicator 1 and 3)	• Mixed-methods approach: ○ first stage: theory-based approach; ○ second stage: estimation of programme effects at micro or local community level; ○ third stage: aggregation of results to programme area, country level; ○ fourth stage: cross-checking using qualitative techniques and structural modelling aiming at triangulation.
2. To what extent has the RDP contributed to employment creation? (Impact indicator 2)	
3. To what extent has the RDP contributed to improving the competitiveness of the agricultural and forestry sector? (Community strategic priority)	
4. To what extent has the RDP accompanied restructuring of the dairy sector? (Health Check objective)	
5. To what extent has the RDP contributed to introduction of innovative approaches? (Health Check objective cutting across axes)	

In relation to measure-level EQs

Evaluation Question	Recommended approach
6. How and to what extent has the measure contributed to improving the competitiveness of the beneficiaries?	• Mixed–methods approach, incl. quasi-experimental design e.g. PSM (GPSM), RDD, IV at micro-level combined with qualitative techniques aiming at triangulation.
7. What other effects, including those related to other objectives/axes, are linked to the implementation of this measure (indirect, positive/negative effects on beneficiaries, non-beneficiaries, local level)?	• Mixed-methods approach: ○ Theory-based and qualitative techniques at the first <orientation> level; ○ Quasi-experimental design, e.g. PSM (GPSM), RDD, IV methods aiming at provision of quantitative results concerning indirect effects (second stage).

4.3.4.4 Suggested methods in relation to Axis 2

In relation to programme-level EQs

Evaluation Question	Recommended approach
8. To what extent has the RDP contributed to protect and enhance natural resources and landscape including, biodiversity and HNV farming and forestry? (Community strategic priority. Biodiversity is also a Health Check objective, Impact indicator 4, 5 and 6)	• Mixed-methods approach: ○ first stage: theory-based approach; ○ second stage: estimation of programme effects at micro or local community level; ○ third stage: aggregation of results to programme area, country level; ○ fourth stage: cross-checking using qualitative techniques aiming at triangulation.
9. To what extent has the RDP contributed to the supply of renewable energy? (Health Check objective cutting across axes, Impact indicator 7)	
10. To what extent has the RDP contributed to climate change mitigation and adaptation? (Health Check objective cutting across axes)	• Mixed-methods approach: ○ first stage: theory-based approach; ○ second stage: estimation of programme effects at micro or local community level; ○ third stage: aggregation of results to programme area, country level; ○ fourth stage: cross-checking using qualitative techniques and/or structural models linked to climate models aiming at triangulation.
11. To what extent has the RDP contributed to improvement of water management (quality, use and quantity)? (Health Check objective cutting across axes)	• Mixed-methods approach: ○ first stage: theory-based approach; ○ second stage: estimation of programme effects at micro or local community level; ○ third stage: aggregation of results to programme area, country level; ○ fourth stage: cross-checking using qualitative techniques and water management models aiming at triangulation.

In relation to measure-level EQs

Evaluation Question	Recommended approach
12. How and to what extent has the measure contributed to improve the environmental situation?	• Mixed-methods approach, ○ incl. theory-based and qualitative techniques at the first <orientation> level; ○ combined with quasi-experimental design, e.g. PSM (GPSM), RDD, IV methods aiming at provision of quantitative results concerning indirect effects (second stage)
13. What other effects, including those related to other objectives/axes, are linked to the implementation of this measure (indirect, positive/negative effects on beneficiaries, non-beneficiaries, local level)?	

4.3.4.5 Suggested methods in relation to Axis 3

In relation to programme-level EQs

Evaluation Question	Recommended approach
14. To what extent has the RDP contributed to improving the quality of life in rural areas and encouraging diversification of the rural economy? (Community strategic priority)	• Mixed-methods approach: ○ first stage: theory-based approach; ○ second stage: estimation of programme effects at micro or local community level; ○ third stage: aggregation of results to programme area, country level; ○ fourth stage: cross-checking using qualitative techniques aiming at triangulation
15. To what extent has the RDP contributed to introduction of innovative approaches? (Health Check objective cutting across axes)	
16. To what extent has the RDP contributed to creation of access to broadband internet (including upgrading) (Health Check objective)	

In relation measure-level EQs

Evaluation Question	Recommended approach
17. How and to what extent has the measure contributed to the economic diversification of the beneficiaries?	• Mixed–methods approach, incl. quasi-experimental design e.g. PSM (GPSM), RDD, IV at micro-level combined with qualitative techniques aiming at triangulation.
18. How and to what extent has the measure contributed to the economic diversification of the beneficiaries?	• Mixed-methods approach, incl. theory-based and qualitative techniques at the first <orientation> level combined with quasi-experimental design, e.g. PSM (GPSM), RDD, IV methods aiming at provision of quantitative results concerning direct effects (second stage).
19. What other effects, including those related to other objectives/axes, are linked to the implementation of this measure (indirect, positive/negative effects on beneficiaries, non-beneficiaries, local level)?	

4.3.4.6 Suggested methods in relation to Axis 4

EQs in relation to Leader

Evaluation Question	Recommended approach
20. To what extent has the RDP contributed to building local capacities for employment and diversification through LEADER? (Community strategic priority)	• Mixed-methods approach: ○ incl. theory-based and qualitative techniques at the first <orientation> level; ○ combined with quasi-experimental design, e.g. PSM, RDD, IV methods aiming at provision of quantitative results concerning direct and indirect effects of LEADER (second stage).
21. To what extent have LAGs contributed to achieving the objectives of the local strategy and the RDP?	• Mixed-methods approach: ○ incl. qualitative techniques at the first <orientation> level; ○ combined with quasi-experimental design, e.g. PSM, RDD, IV methods aiming at provision of quantitative information about effectiveness of LAGs (second stage).
22. To what extent has the Leader approach been implemented?	• Qualitative techniques
23. To what extent has the implementation of the Leader approach contributed to improving local governance? (Community strategic priority)	• Mixed-methods approach: ○ incl. qualitative techniques at the first <orientation> level; ○ combined with quasi-experimental design, e.g. PSM, RDD, IV methods aiming at provision of quantitative information about effectiveness of Leader approach (second stage).

4.3.4.7 Suggested methods for the assessment of TA and NRN

EQs in relation to TA and NRN

Evaluation Question	Recommended approach
24. To what extent has the NRN contributed to RDP objectives?	• Mixed-methods approach: ○ incl. qualitative techniques at the first <orientation> level; ○ combined with quasi-experimental design, e.g. PSM, RDD, IV methods aiming at provision of quantitative information about effectiveness of NRN (second stage).
25. To what extent has the TA contributed to RDP objectives?	• Mixed-methods approach: ○ incl. qualitative techniques at the first <orientation> level; ○ combined with quantitative and qualitative methods using as inputs data concerning programme effectiveness and efficiency (results of evaluation under Axis 1, 2, and 3).

4.4 Do's and don'ts in relation to the assessment of evaluation methodologies (based on experience)

DOs	DON'Ts
Try to assess programme effects (results/impacts) on the base of appropriate counterfactual. Counterfactuals can be constructed using various methodologies, incl. qualitative methods.	Use approaches with no reference to counterfactuals
Apply to evaluation a mixed methodological approach. The best empirical approach is a combination of rigorous quantitative and qualitative methodologies.	Apply simplistic or "naïve" techniques to estimate programme effects
If using qualitative methodologies apply the broad range to answer the "why and how questions": • Why does the programme work (or not work)? • How does the programme achieve its goals? • Why does it work for some clients (or in some situations) and not for others? • What are/were the needs of the clients that were not anticipated by programme developers? • What were the additional unintended and/or unexpected positive or negative consequences?	
Stakeholders involved in focus groups should be representative for a given type of stakeholder (key stakeholders)	Involve too many stakeholders in focus groups.

DOs	DON'Ts
Remember that application of different methods, even using the same data set, may lead to different results, incl. the estimated magnitude of programme effects and even its sign (+/-).	Leave out unverified possibility to undertake more surveys in order to collect more data, especially in a situation when expected trade-off between data availability and introduced bias and systematic errors is large.
Remember that methodologies selected to construct counterfactuals should meet widely acceptable evaluation standards and basic criteria used for the judgement of evaluation's quality.	Start collection of data on programme non-beneficiaries after the end of the programme only.
As not all evaluation techniques are substitutable with each other, choose an evaluation design based not on a single method but on a bundle of methodological approaches which are internally consistent and supplementary to each other.	Select only one method (e.g. theory based) and use it across all axis Forget that there is no single method satisfying all evaluation requirements.
Remember, that a deliberate decision should be made in choosing an appropriate bundle of methods in order to ensure a minimisation of potential bias and systematic errors in produced results on one side, and the highest possible evaluation quality on the other side.	Leave out a possibility to examine credibility of your selected methods by comparing them to the best benchmark.
Remember that a decision to use or not to use certain evaluation design (especially those which are based on counterfactuals), and instead to apply less rigorous or simplistic techniques, only because data availability may be restricted, may be highly distortive and may lead to producing unrealistic assessments of programme results/impacts.	Rely too much on personal interviews with key stakeholders without cross-checking and appropriate filtering obtained information.
One should rather consider carrying out additional surveys in order to collect additional necessary data on characteristics and performance of programme beneficiaries and non-beneficiaries instead of taking a risk of overstretching the interpretation of programme results based on simple yet biased evaluation techniques.	

5. DATA

In all *ex post* evaluations of RDPs good data availability and data quality will be essential for appropriate assessment of the programme effects. In general, adequate data should be collected for pre-defined RDP potential beneficiaries (farmers, foresters, businesses, municipalities, etc.) as well as for relevant comparison groups in order to establish appropriate counterfactuals (e.g. non-beneficiaries farmers). Collected data should: i) enable exploring heterogeneous effects; ii) have sufficiently large sample size to achieve statistical power; and iii) contain a full range of variables to be used as controls.

5.1. Major challenges in data management and data collection during the evaluation of the RDP

First, it is advisable to explore in advance which data already exists and what still needs to be collected. Using existing primary and secondary data (existing data within MA, PA, implementing bodies, various research institutions, regional and local authorities, statistical offices, NGOs, government agencies, etc.) may save significant amounts of resources. However, it is important that the collected data is available in the required quantity and quality for robust evaluation in terms of timing, completeness, adequacy and reduction of potential bias.

Second, if new data has to be collected a number of additional concerns need to be addressed: timing, sample design and selection of appropriate survey instruments. Adequate data should be collected in due time and in a cost-effective way while respecting evaluation requirements. The cost-effectiveness is here an issue, since the collection of additional data might represent an important fraction of total costs of evaluation (may easily represent ¼ to ⅓ of the total costs of an evaluation). In order to be effective, careful planning of data collection modalities in advance should be initiated already at the beginning of programme implementation.

Utilised data can have different forms: cross-sectional data[153], time series data or panel data. A panel dataset (or longitudinal data) consists of a time series for each cross-sectional member in the dataset. While panel data requires replication of the same units over time (ideally prior to and after the implementation of a given RD programme), panel datasets, especially those on individual farms, enterprises or holdings are not easy to obtain. Yet, observing the same units over time leads to several advantages over other forms of data sets (e.g. cross-sectional data in one time period). One of the benefits is that having multiple observations on the same units allows the analyst to control certain unobserved characteristics of individuals or households (or the unit of analysis). Due to these advantages, economists and evaluators now recognise that many of evaluation questions are difficult, if not impossible, to answer to satisfactorily without panel data[154].

Steps in data collection for evaluation

Collection of micro- and regional level data necessary for evaluation of RDPs based on counterfactual evaluation design should be carried out as much as possible in a standardised manner.

The main organisational steps in collecting of data could be as follows:

- Prepare a list of economic, environmental and other specific micro-data is important to be collected on programme beneficiaries and non-beneficiaries to conduct the evaluation, by type of beneficiary for all programme measures:

[153] Data covering separate units in a given time period.
[154] Comp. Wooldridge 2002; World Bank, 2011b

  - Farms
  - Food processors
  - Forestry enterprises
  - Micro-level enterprises
  - Municipalities
- Prepare a list of regional data to be collected;
- Prepare a description of data availability for a given country and RDP territory;
- Prepare a list of data owners (institutions, organisations, networks, etc.) to be approached in order to collect this data;
- Describe actions needed to obtain existing data that is important for evaluation,
- Take into consideration necessary institutional arrangements to ensure confidentiality and anonymity of the collected data (including contracts, agreements, legal procedures, etc.);
- Describe a form of data collection and its delivery to the evaluator.

5.2. Which data is necessary for the evaluation of RDP results at beneficiary or micro level?

Economic results

Counterfactual design methods are relatively data demanding. Collected economic data should include all relevant information on programme beneficiaries and non-beneficiaries regarding their structure and performance and should cover periods “before” and “after” the implementation of the programme. The biggest part of collected data (approx. 80%) is usually related to the data block “structure” and is used to construct meaningful control groups (e.g. via the application of matching techniques, etc.). Then most of the collected result and impact indicators are part of the data block “performance”. Depending on the level of analysis, required data should either describe micro or regional or national levels structure and performance of analysed units (in both cases collected data should concern programme beneficiaries and non-beneficiaries). The micro-level data may consist of bookkeeping data or survey data (or both). Collected data should clearly identify programme beneficiaries and the level of support they received from individual RD measures (e.g. 121, etc.). Ideally, micro-level data panel should comprise no less than 150 beneficiaries enterprises, farms or holdings for each of analysed measure (or a group of measures) and 2-3 times more non-beneficiaries.

Data at micro level should be collected from beneficiaries and non-beneficiaries on the basis of secondary data and/or own surveys. The FADN database provides data of beneficiaries and non-beneficiaries for agricultural RD measures as: farm investment support (measure 121), LFA schemes (measures 211, 212) and agro-environmental measures (AEM, e.g. measures 214, 215). The suitability of FADN data for RD evaluation could be enhanced if different types of RD support could be further disaggregated and/or separately listed (e.g. differentiation of various AEM or investment types). Outside the agricultural sector, the availability of secondary data becomes scarce. Various national statistics on households, municipalities, small businesses are available but their accessibility varies strongly among Member States. In many cases, own surveys will be the only source of information for the evaluation of non-agricultural RD activities.

5.3. Which data is necessary for the evaluation of RDP results and impacts at RDP and national level?

Economic results and/or impacts

Data for macro-level analysis (national level) are to be collected from official statistics. NUTS 3 – NUTS 5 might be the lowest possible macro level where effects of the RD support can be estimated (although this may vary according to the Member States). For example, in case of employment, relevant data may stem from regional economic accounts including data on employment (number of jobs or FTE, sometimes differentiated by sectors). Additional information on NUTS 3 level (e.g. wages, population size, firm structure, etc.) should be available from various official statistics. Further data on labour statistics are sometimes provided by national labour offices. They typically count employed persons excluding self-employed and civil servants. The data distinguish between full-time, part-time and marginally employed persons. However, this data source does not help if the majority of employment effects of the RD support are expected to impact the number of self-employed persons (farmers, start-ups in diversified enterprise, etc.). Additional surveys or case studies will therefore be required to get the necessary information on employment.

In case of RDP and national analysis, counterfactual design requires data to be collected for individual regions (e.g. NUTS 4 – NUTS 3), communities (e.g. NUTS 5) or villages. The results generated at these levels will be thereafter aggregated at RDP or national level. In the majority of cases detailed (secondary) data on regions (NUTS 3-NUTS 5) can only be obtained from respective statistical offices. Data on small communities or villages may be collected through surveys.

Data on RD support at RDP and national level can be collected as follows:

- From the paying agencies at micro level broken down by type of supported activity, type of supported beneficiary, regional or postal zip code of the supported project, public expenditures, etc. Aggregation of total expenditures at NUTS 3 or 2 levels is possible via regional codes and/or postal zip codes. Support provided by other instruments (1st pillar policies, EU Structural Funds, etc.) should also be taken into account, especially when fitting macro level models.
- Sometimes data of RD support at a regional or programming area is available directly from the PA, and on support from other instruments, e.g. Structural Funds - from respective national agencies.

Environmental results and/or impacts

Data on reversing biodiversity decline: *population of farmland birds*

At present, data for the calculation of Farm Bird Index (FBI) originates from national monitoring of widespread birds collected and compiled by the Pan-European Common Bird Monitoring Scheme (PECBMS) in cooperation with Statistics Netherlands. In 2013, 40 European countries were directly involved in the collection of data for the index: Andora, Austria, Belarus, Belgium, Bosnia-Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Macedonia, Germany, Hungary, Iceland, Ireland, Italy, Latvia, Lichtenstein, Lithuania, Luxembourg, Malta, Moldova, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine and United Kingdom. Information on national monitoring schemes, indicators and the use of their results can be found on the EBCC website[155].

Most EU countries have well established and high quality common bird monitoring schemes running – by relying and building on these existing schemes, Member States can greatly reduce the cost and administrative effort of collecting, processing and interpreting bird data for the purpose of rural

[155] European Bird Census Council.

development monitoring and evaluation. Although national schemes differ in count methods in the field, these differences do not influence the supranational results because the indices are standardised before being combined.

Data from generic monitoring schemes may often not be ideal for specific evaluation purposes, either because of poor geographical coverage of an area, or because they do not provide highly detailed site specific information on numbers and distribution required for these purposes. Therefore it is vital for the evaluator in the first instance to ask local monitoring data providers to verify the representativeness of their data, especially from farmland under agro-environmental measures point of view. If it is clear that monitoring data does not represent the target area it should be clearly mentioned in the report and linking AEM with FBI should be made very cautiously.

Sustainability of data collection: though the actual common birds' surveys are carried out mostly by volunteers and thus represent a unique low cost resource of data, some funding is still needed for coordination and support to the volunteers' network in getting suitable form of data which fits the RDP evaluation and for the data processing phase.

The reliability of the data: it is important to have a large and representative set of sample points and species – in the case of an indicator based on multispecies index of changes in abundance, in general the more species contributing to the indicator, the more reliable.

Only good indicator species should be included (e.g. farmland specialists likely to be affected by changes to habitat quality, species easily detected in common bird surveys), remembering, however, that some difficult species to find may be the most vulnerable. Individually, many species may show annual changes in abundance that may reflect a variety of environmental factors, such as extreme weather conditions during the breeding season, poor conditions on the winter grounds, changes in predation pressure, and simple sampling error and statistical noise. Long-term monitoring and use of a wider range of species helps detecting the underlying trends. The evaluator's expertise is essential for isolating the effects of the measures from other contextual factors.

To guarantee the quality of the national FBI it is necessary to have an appropriate national monitoring scheme. Often national monitoring schemes do not cover representative amounts of farmland and therefore are weak in assessing real changes in populations of farmland birds. FBI is presented as one number without explanations. So, no-one knows what is behind the values submitted by the countries and what aspects should be taken into account for each of them. For example, in Estonia a lot of FBI monitoring points are located in forests and not on agricultural land. Such problems could be overcome by the species selection – indicator species should show a strong linkage to farmed land. In addition, many FBI monitoring points are located in protected areas. Even the proximity of protected areas to agro-environmental schemes bird impact indicator monitoring places can affect the results: protected areas are favouring high number and species richness of birds which may disperse to agricultural land in the vicinity. In such agricultural lands the results of bird population are overestimated. The possibility to use an alternative composition of bird species where this is appropriate to national or regional situation – while FBI calculates the aggregated EU-wide indicator on a basis of a "basket" of species tailored to best capture EU-wide trends, its use at national or regional level requires "species baskets" tailored to the local conditions (i.e. including species that are good indicators for farmland habitats and are common enough to be captured in common birds surveys). So MS may use an alternative composition of bird species (can of course include more or less species than the EU-wide one) where this is appropriate to national or regional situation.

Data on reversing biodiversity decline: *area under successful land management contributing to biodiversity*

If monitoring aims to evaluate changes over time (either within measure sites, or relative differences between measure and control sites), then evaluations must take place over an appropriate length of

time (dependent on species or community being evaluated). Such evaluation may be of the greatest value when comparisons can be made against 'baseline data' (data collected prior to the implementation of the measure).The collection of data should enable the comparison between sites where the measure is being implemented and control sites (i.e. otherwise similar sites where the measure is not being implemented,). If the species or community to be monitored is known to be scarce or range restricted, then selecting survey sites entirely at random may be undesirable, as many survey units will be outside the species range. In order to sample more effectively, stratification can be used to target areas within the known range. Survey sites (measure and controls) are then picked at random within delimited areas (or set percentages of survey sites are selected to fall within and out of the delimited areas). Details of site location, crops, managements, habitats, landscape features, etc. within and around the survey units should be collected for use as controls in the data analysis, to account for variation not directly explained by the measure.

However the following challenges for the data collection may occur:

There are limits on previous or ongoing monitoring data available on birds, and funding challenges for carrying out special field studies. One has to bear in mind that the EBCC Pan-European FBI measures biodiversity at a continent or subcontinent level, whereas the need to measure the impact of RDP is a national or regional challenge in each MS. Therefore it is the obligation of the MS to care for the appropriate funding of this data collection. There might also be a need for cross-regional training, not in fieldwork as such but in setting down appropriate and most cost-efficient national schemes, and processing the results in a correct and relevant way. Lack of monitoring data and finances to carry out special studies may lead to inadequate or misleading evaluation results.

The suitability of previous or ongoing monitoring data for evaluation of bird impact indicator – in the case of using some previous or ongoing monitoring data for the evaluation of bird impact indicator the background and purpose of the data collected needs to be assessed to decide the suitability of the data. E.g. one problem could be that often areas with good environmental state are preferred – so, areas with not so good environmental state are not monitored. If such data is used, a wrong picture about the overall situation in the countryside might be reflected.

Formation of monitoring sample – the main weakness of most monitoring programmes is not the actual counting method but the sampling strategy. The formation of monitoring sample is a challenging task. There are no identical farms – the landscape is different – the share of semi-natural habitats is different – farm sizes are different, etc. All these differences that could affect the bird impact indicator results should be taken into account while forming the monitoring sample. Samples have to be representative in terms of habitat and geographical location. Therefore the best way to gather a representative sample is to use a large number of random plots[156].

Comparison group – There should also be a comparison group in the sample to find out what is the situation without the support; thus areas without the support should be involved. In some MS, e.g. Austria, Finland and Sweden the formation of comparison group for agro-environmental measures is practically impossible as they cover 90% and even more of the agricultural land. In such cases, monitoring results of different measures (e.g. broad and shallow measures versus specific species or habitat targeted measures) together with the analysis of differences of measure requirements could form the basis for comparison.

Data collection methodology – There are several methods for bird data collection with a different quality and in different circumstances. It is important to use standardised methods as much as possible, so that the same method is always used on the same site every year. Results that come

[156](http://www.ebcc.info/wpimages/other/14-SVENSSONH.pdf)

from a mixture of methods are equally as useful as results from a single method as long as a standard is maintained through time for each individual count.

Data on HNV areas

In most regions, data limitations mean that using species distribution alone is not a sound approach to identifying HNV farmland and forestry. The available data is not sufficiently exhaustive, either in terms of the range of species covered, or of geographical coverage, and is not updated with sufficient regularity.

Hence, species distribution data may be useful for filling gaps left by the rather crude approaches currently possible using land cover and farming systems data. These are likely to miss some areas of farmland of high nature value, especially those characterised by smaller semi-natural elements and mosaic patterns, or farmland under more intensive use that continues to harbour species of conservation concern. Species data is particularly relevant for identifying Type 3 HNV farmland.

The JRC/EEA has applied this approach. Lists of indicator species for HNV farmland have been defined and corresponding sites from Important Bird Areas and Prime Butterfly Areas have been selected and mapped, to complement the land cover picture. Several countries have used national species data for this purpose in their HNV farming identification exercises, for example Bulgaria, Romania, Sweden and England. However, this approach does not, on its own, make a link between biodiversity and farming systems or practices. Once areas have been identified on the basis of species data, it will be necessary to establish what land cover patterns and farming systems or practices are associated with the high level of biodiversity, in order to monitor changes in these characteristics.

Current EU, national and regional databases were not designed for monitoring tendencies in HNV farming. At best, they only provide time-series data on some of the HNV variables. Data collected at regional or national level therefore would need to be complemented by more detailed sample surveys.

Currently for most regions, sample surveys are the only way to capture tendencies in certain variables, which may in fact be the most important ones, especially tendencies in:

- The intensity of use on the farmed or forest area, and specific practices;
- Condition of semi-natural land cover at farm or forest level – semi-natural pastures, hay-meadows, orchards, woodlands, and smaller semi-natural landscape features such as hedges and ponds.

Surveys of established sample sites could be complemented with random sample surveys outside these sites to construct counterfactuals. Random sample surveys of farming practices are undertaken as part of FSS data gathering, and could be extended to cover HNV farming criteria.

Further development of existing databases is an important consideration for the future of HNV farmland and forestry monitoring. It would be desirable to incorporate HNV variables in existing databases, especially in agricultural census database (FSS, Eurostat) and identification system for agricultural parcels (LPIS-IACS), including:

- Parcels consisting of semi-natural farmland, including traditional orchards and hay meadows, and smaller features such as hedges and ponds;
- Common grazing land used by the farm (area used in ha or LU grazing days);
- All forage land used by the farm (including scrubby and woody forage);
- All grazing livestock present on the farm.

First steps towards testing the incorporation of these data could be taken for the sample survey sites, especially for LPIS.

Data on water quality

The mass balance is rather difficult to calculate as it requires the real N input into the farm. In order to apply the suggested "difference in difference" approach detailed information on nutrient amounts used on farms or micro-level is required. In Germany, for example, such data should in principle be available on farms since farmers are supposed to calculate so-called annual and multiyear nutrient management plans to record the farm nutrient inputs and outputs. However, while beneficiaries of RD measures should provide the necessary data for this exercise, this cannot be expected from non-beneficiaries.

At the regional level, the percentage of different land cover, land use and farming types are the key data required to evaluate the gross N balance together with atmospheric N fixation and deposition. For the quantification of land use and land cover a variety of data sources is available. Besides the farm structure survey (FSS) and IACS data land cover can be determined due to the development of remote sensing devices, digital elevation models and GIS software. All European countries have access to these tools and are currently using them. Today there are about 20 years of data sets, which allow analysing land use and cover changes with high spatiotemporal accuracy. Data on livestock is also provided by the FSS and livestock surveys.

Yet, the main issue with these models is to translate land cover into land use (e.g. from grassland land cover to fertilised, heavily grazed or hay production land use) in order to apply the most relevant coefficient of N input and output[157] and to harmonise them within the European Union. This will not solve the problem of the lack of evaluation of the results obtained but at least it will allow for comparison of the results obtained.

The gross nutrient balance[158] (GNB) represents a simplification of complex interrelated and variable processes. The calculation of GNB comprises uncertainties that are mainly based on the statistical database or the coefficients used to convert statistical data into nutrient quantities (e.g. livestock into organic nutrients). Specifically, if the amount of mineral fertilisers can be reasonably determined, the N application in manure is not obvious since it depends on the type of manure, the storage method and retention time before application. The second problem is related to the fact that N balance is the result of transfer of N mass from one element of the dynamic farming system (e.g. fertiliser and manure) to the other (e.g. crop, crop residue, fodder, and animal production). This transfer of mass is accounted for by using conversion coefficients. Schroder et al. (2003) identify 4 main conversion coefficients: SH to convert soil-N into harvestable crop-N; HF to convert harvestable crop-N into feed-N; FP to convert feed-N into animal produce-N and MS to limit the gaseous losses from stables, manure storages and manure application.

The main challenge for the future is to harmonise these conversion coefficients within the EU Member States in order to compare the results. It entails the publication of very different results from different countries or regions. For instance nitrogen manure coefficients from northern European countries varied by 80% for heifer 100% for dairy cow, and 400% for pig[159]. This harmonisation is in particular essential for comparing GNB levels. However, the issue is of less importance for the comparison of GNB changes as required by CMEF.

The gross nitrogen balance represents the theoretical nitrogen surplus in the soil calculated by the difference between the total quantity of nitrogen inputs entering the soil and the quantity of nitrogen outputs leaving the soil annually. The use of gross nitrogen balance as an indicator of the potential N

[157]Verburg, P.H., van de Steeg, J., Veldkamp, A., Willemen, L. (2009): From land cover change to land function dynamics: A major challenge to improve land characterisation. Journal of Environmental Management, 90: 1327-1335
[158]http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/en/aei_pr_gnb_esms.htm
[159]vanEerdt, M.M., Fong, P.K.N. (1998): The monitoring of nitrogen surpluses from agriculture. Environmental Pollution, 102: 227-233

loss to aquatic system is significant. Gross nitrogen balance does not inform on the form (organic, ammonia, nitrate) in which nitrogen is in the soil. If nitrate is the form much more prone to leaching, organic N is rather stable and is function of the carbon concentration in the soil. A better evaluation of the N risk to water quality would require estimation or measurement of gas emission (Net Nitrogen Balance). Ideally, water quality monitoring (nitrogen fluxes measurements at the outlet of agricultural catchments) would be the best method.

Aggregating farming region gross nitrogen balances at the country scale does pose specific limitations in information because it masks the heterogeneous reality of responses expected from different regions due to inherent differences such as geology, soil, climate and socio-economic context. Aggregating scores by country, if necessary, should be a function of the rate of change of local or regional agricultural areas. Therefore, the results should be expressed as a rate of change of estimated N surplus by "representative region" (NUTS3 or smaller). This approach would provide several useful indicators to quantify the impact of rural policy on water quality. It would allow to: i) measure the impact of rural policy on water quality in different agricultural contexts; ii) determine the most "receptive" contexts where improvement is noticed and the least "receptive"; iii) analyse the reasons of such differences at the appropriate scale; iv) compare the rate of change in "similar" agricultural regions among countries.

The gap between gross nitrogen balance measurement and the impact on water and air quality could be easily filled, i) by measuring water quality in agricultural catchment; and ii) by using net nitrogen balance implying to estimate N gas emissions. This would allow developing a real systemic appraisal of the consequences of the agricultural policies and practices.

In water quality measurement, it has to be emphasised that the assessment of impacts of RD measures in the field of water quality by using GNB should provide foremost information on nitrogen leaching or evaporation risk but not of direct water/air pollution. It is therefore a tool to evaluate the potential impact of RD measures on the environment – in particular water. It should show whether RD measures increase or decrease the risk of water/air pollution. – Thus when assessing this indicator, the currently available information (e.g. from WFD) should be adapted to the necessities for capturing the impacts. The measurement of water quality parameters, including all the nitrogen species in streams, lakes and groundwater is the most accurate method to assess water quality and its evolution. This is the basis of the EU Water Framework Directive (WFD). There is already an important network of monitoring sites throughout Europe with rather long-term data sets (usually up to 30 years of data). The primary goal of this WFD monitoring is to assess the water quality of streams, lakes and groundwater and follow its evolution. The data set gathered is compared against water quality thresholds which allow classifying water bodies.

Combating climate change

Specific RDP evaluation experience on climate change is limited, but the measures above are dominated by those involving field cultivation (area/yields). This suggests that existing methods for determining area and yield additionally are applicable for energy crops.

Methodologies include:

- Quantitative methods: Farm scale modelling scenarios, 'DiD' method and/or Propensity score matching approach (participation or land conversion under programme conditions relative to counterfactual).
- Qualitative methods: farm interviews, reflective comparisons for farmers who are growing crops or who have installed capacity in biogas, wind or hydro.

Data sources for all sources are relatively good:

- Farm scale data on land under specific crops (FADN),
- Data can be partitioned in a variety of ways e.g. before/after programme; farm type and size,
- Information on installed capacity (biogas and wind),
- Conversion factors to express biomass, gas or wind in terms of tonnes of oil equivalent,

For wind or biogas generation, a distinction is made between energy use on farm and whether energy is exported to a national grid. For the latter the amount of renewable energy generated is based on: (a) Metering; (b) calculations based on installed capacity and (c) average load factors; where there aren't meters or a verified installed capacity. To determine this information, farmers and other land managers would be required to submit information of compliance with specific measures. This information could be subject to surveillance using systematic or randomised farm surveys and/or possible forms of triangulation using GIS or satellite data on land use. The same mitigation evaluation procedure could be incorporated into the evaluation procedure that is currently used for agro-environmental schemes.

Social results and/or impacts

Secondary data to be used in measuring social results and impacts should be found in national and regional statistics. With respect to beneficiaries, data should be available from the monitoring conducted by the Managing Authority, paying agencies, implementing bodies and LAGs, local development strategy, monitoring databases based on project applications, business plans, project contracts and requests for payments.

In relation to LAGs they were not asked to conduct systematic evaluation activities in the programming period of 2007-2013. This has caused large data gaps in the evaluation of social results and impacts with respect to Leader. This was partly bridged with various evaluation studies, in which Managing Authorities and evaluators have collected information to assess RDP social results and impacts. These studies can be used as sources of information in the *ex post* evaluation. In addition data and information collected through self-assessment activities[160], conducted voluntarily by LAGs, sometimes harmonised across the entire region or even RDP area, including defined observation fields, indicators and rating criteria that can be used as additional sources of information by the *ex post* evaluator.

Assessing social results and impact data and information can be gathered though personal observations and interviews, i.e. starting with online surveys up to individual or grouped in-depth interviews. The information raised there is partly quantifiable (e.g. standard questionnaires featuring a limited number of pre-established answers; or rating scales), partly usable as narratives in order to explore cause-effect relationships.

When designing questions with rating scales, it should be made clear

- What is actually rated (the subject)
- What do the lowest and highest possible points mean.

[160] In the Austrian case, self-evaluation of LAGs was based on a country-wide set of eight indicators: Coordination processes, national and trans-national cooperation, distribution of projects across fields of action, marketing and public relations, participation in the Leader network, balanced involvement or men and women, youth involvement, inter-municipal participation and co-operations. These indicators were rated annually using quality scores from 0 to 5. Source: Dax T., Oedl-Wieser T., Strahl W. (2011): Halbzeitbewertung der LEADER-Maßnahmen im Rahmen des ÖPELR 2007-2013.BundesanstaltfürBergbauernfragen, Wien.

[160] *„LAGs should be trained in self-evaluation by evaluation specialists."* Source: DG AGRI Guide (2011) for the Application of the Leader Axis of the Rural Development Programmes 2007-2013 funded by the EAFRD. Bruxelles, p. 24.

There are basically two types of rating scales:

- Open scales with just numbers (e.g. from 1 to 7) to estimate the relative quality of something. In many cases, it is advisable to use an impair scale with scores ranging up to 5 or 7, in order to keep the interviewees from converging towards the "neutral" middle: impair scales do not have this.
- Graded scales featuring a description of the state of affairs according to each level. For instance, such scales are in use for the voluntary self-evaluation system in Austria. They follow a balanced scorecard format with four main perspectives: (i) Resources, (ii) Learning, (ii) Results and Impacts, and (iv) Internal and implementation processes.

Narratives are necessary to come to policy recommendations which go beyond a bony answer to the question on whether a measure or project has been successful or not. The narratives provide the flesh of a policy recommendation. Obviously, the evaluation team feels seduced to import and impose its proper preconceptions, basically when acting alone, without critical counterparts. Everybody can succumb to the constant seduction to find simple answers in a complex world. Therefore the evaluation design should be made in a way that multi-perspective be granted during all stages until drafting the final report.

5.4. Specificities in relation to TA and NRN

Technical assistance

Monitoring data, collected by paying agencies, implementing bodies and Managing Authorities should be used in the evaluation of technical assistance as the source of quantitative evidence. In addition the qualitative information will be needed in order to assess technical assistance results and impacts. The qualitative data will be collected through interviews, focus groups and surveys with rural development stakeholders who are participating in or benefiting from the activities financed out of technical assistance budget.

National Rural Networks

NRN enjoy certain operative freedom in the definition of indicators and collection of data. Data collection relates to decisions on which methods will be applied in their assessment and what budget will be allocated for that purpose. Cost-efficiency is often the factor influencing the selection of methods, sometimes to the disadvantage of the quality for evaluation use.

The available data on NRN action plan implementation or NRNP activities are usually properly kept at output level (number of events, number of participants, newsletters produced, etc.) through the monitoring systems.

Data for NRN results (and impact) indicators has to be collected on an *ad hoc* base (e.g. tracking of satisfaction and uptake of good practices by event participants over the programming period) or will almost exclusively have to be based on qualitative data collection for the purpose of the *ex post*. The issues addressed in relation to other RD interventions also apply here too (lack of baseline, time-sensitivity of memory, perceptions sensibility, broad objectives and intangible outcomes).

Collection of information for NRN results and impacts should be done where NRN effects are expected:

- At the level of individuals (micro-level) participating in the network activities, where they acquire knowledge and information in order to conduct their work more effectively (related to multilevel sector governance in the narrower sense);

- At the level of NRN groups of stakeholders and their organisations (meso-level), where effects in relation to building capacities to achieve RDP objectives can be expected, and;
- At the level of rural areas (macro-level), where effects on the improvement of the governance in the rural areas in a broader sense can be expected (e.g. local or territorial governance accommodating property of the decentralisation).

The types of effects above indicate the types of information collected using various methods for evaluation of NRN results and impacts, namely:

- data collected in relation to participation and satisfaction, e.g. via assessment sheets, questionnaires, etc. directly or almost directly at network activities;
- data collected in relation to opinions and connections e.g. via interviews, surveys, self-assessment tools, focus groups, etc. in the course of the routine work of stakeholders and involved actors;
- data collected in relation to governance in the broader sense, e.g. either through social network analysis, visualising the change in the network relations, or through indirect evidence, e.g. through an increase in cooperation becoming manifest in a new institution or memorandum of understanding induced by network activities.

In a schematic way the following data sources can be defined:

- Quantitative data for input, output and result indicators (monitoring);
- Quantitative and qualitative data for result indicators collected from NRN beneficiaries (surveys using questionnaires, interviews, focus groups or case studies, etc.);
- Quantitative and qualitative data for impact indicators collected from both beneficiaries and non-beneficiaries of the NRN in order to conduct counterfactual analysis (surveys using questionnaires, interviews, focus groups or case studies, etc.);
- Official statistics (if available).

It goes almost without saying that these three archetypal data collection approaches and tools behave differently in terms of timing, clarity and suitability. The table below draws a link to the evaluation framework:

Table 11 Data collection approach and tools suitability for the intervention logic level

Data collection approach and method.	Intervention logic level		
	Output	**Result**	**Impact**
Data collection related to participation and satisfaction.	Highly suitable.	Partly suitable (longitudinal surveys).	Not suitable.
Data collection related to opinions and self-assessment.	Partly suitable (highlighting durability of perceptions).	Highly suitable (especially regarding connections and ties to other members of the network).	Partly suitable (reflection and triangulation on data collected by the third type below).
Data collection related to governance in the broader sense.	Not suitable.	Partly suitable (analysis of change in network connections density).	Highly suitable.

5.5. "Dos and don´ts" in relation to using specific data in answering evaluation questions (based on experiences)

DOs	DON'Ts
Screen all existing data sources in relation to the assessment of RDP results and impacts against common and programme-specific indicators.	Decide on data collection before screening the data availability.
Check the quantity, quality and relevance of existing data sources and make arrangements to ensure their availability (institutional, legal, and financial) in the format needed for the assessment of RDP results and impacts.	Use existing data without ensuring their format needed for the RDP evaluation.
Decide on approaches to fill data gaps including those needed for counterfactual and exploring heterogeneous impacts.	Use only existing data on beneficiaries.
Always collect data on programme beneficiaries and appropriate comparison groups (e.g. non-beneficiaries). If "non-beneficiaries" do not exist, collect data on programme intensity levels of individual programme beneficiaries only and apply more advanced counterfactual approaches (e.g. GPSM).	
Remember that collected data should be robust enough to eliminate a possible selection and other forms of evaluation biases.	

PART III: TOOLBOX

1 OBJECTIVE TREES AS PART OF THE RDP INTERVENTION LOGIC BY AXIS

Figure 14 Intervention logic - RDPs 2007-2013

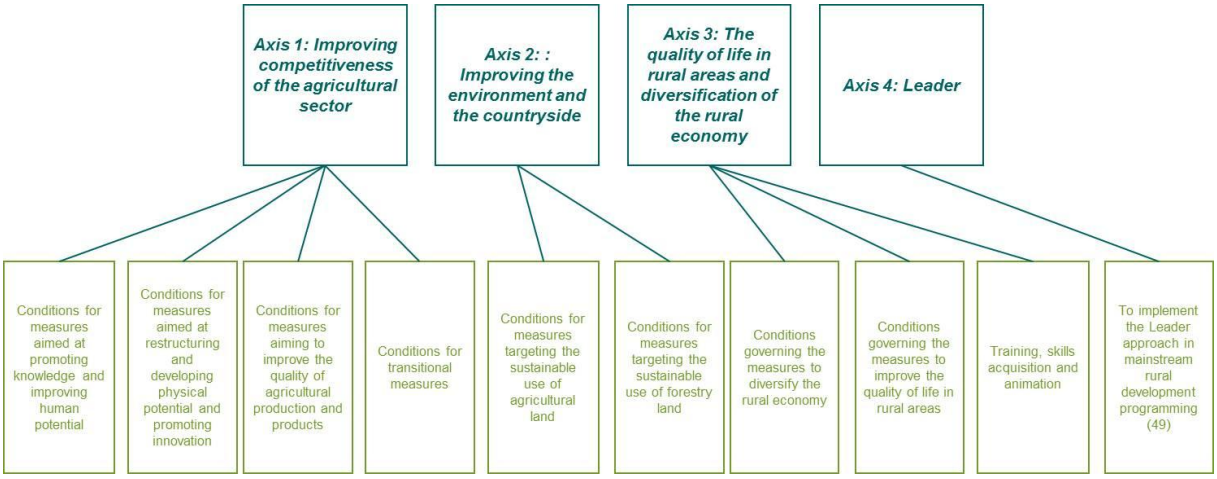


Figure 15 Intervention logic – Axis 1

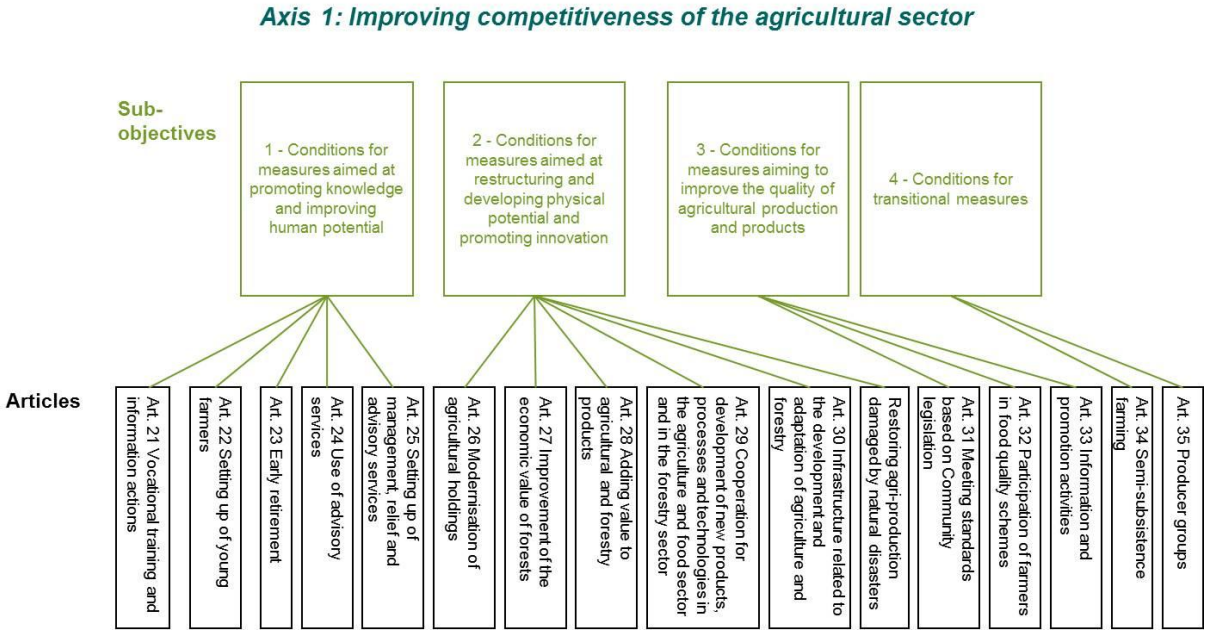


Figure 16 Intervention logic – Axis 2

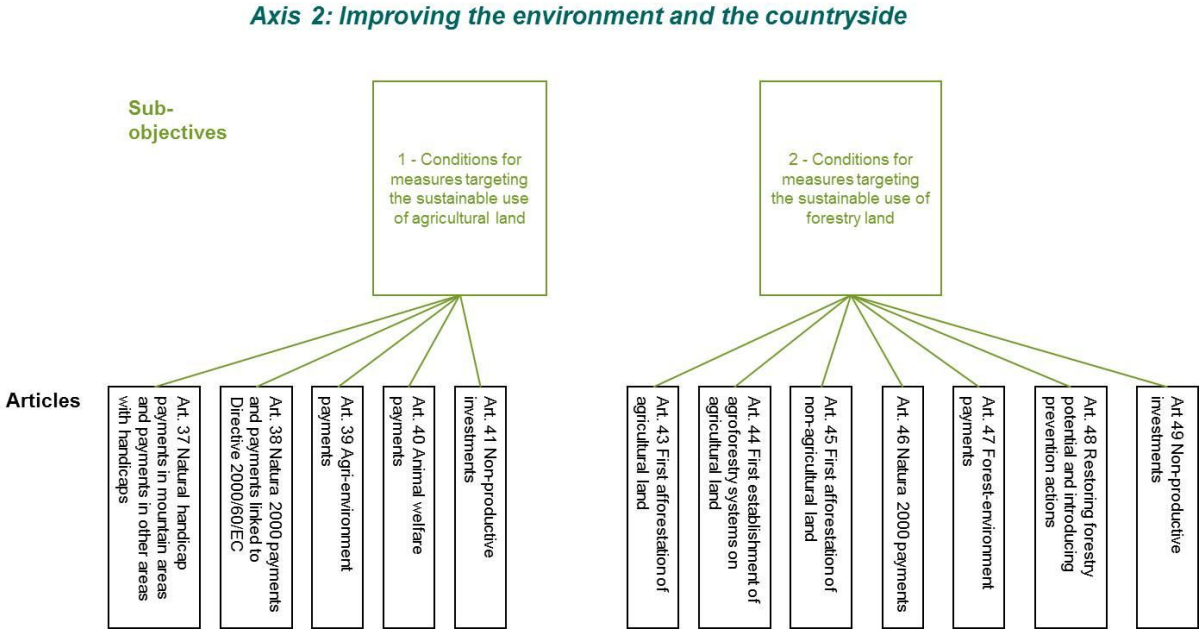


Figure 17 Intervention logic – Axis 3

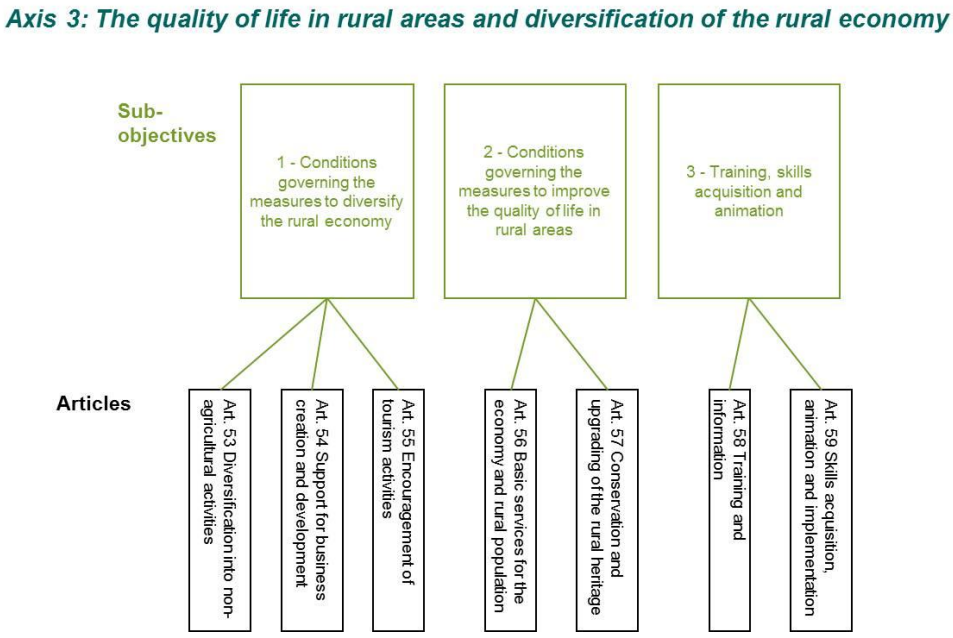
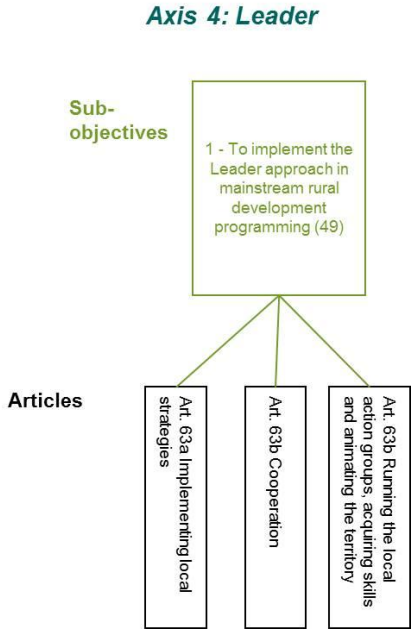


Figure 18 Intervention logic – Axis 4



2 SET OF REVISED COMMON EVALUATION QUESTIONS

Programme-related CEQs

1. To what extent has the RDP contributed to the growth of the whole rural economy? *(Lisbon objective; related impact indicators 1:Economic growth and 3:Labour productivity)*
2. To what extent has the RDP contributed to employment creation? *(Lisbon objective; related impact indicator 2: Employment creation)*
3. To what extent has the RDP contributed to protect and enhance natural resources and landscape including, biodiversity and HNV farming and forestry? *(Community strategic priority, Biodiversity is also a Health Check objective; related impact indicators 4:Farmland Bird Index, 5: High Nature Value Farming and Forestry and 6: Water quality)*
4. To what extent has the RDP contributed to the supply of renewable energy? *(Health Check objective; related impact indicator 7:increase in production of renewable energy)*
5. To what extent has the RDP contributed to improving the competitiveness of the agricultural and forestry sector? *(Community strategic priority)*
6. To what extent has the RDP accompanied restructuring of the dairy sector? *(Health Check objective)*
7. To what extent has the RDP contributed to climate change mitigation and adaptation? *(Health Check objective)*
8. To what extent has the RDP contributed to improvement of water management (quality, use and quantity)? *(Health Check objective)*
9. To what extent has the RDP contributed to improving the quality of life in rural areas and encouraging diversification of the rural economy*? (Community strategic priority)*
10. To what extent has the RDP contributed to introduction of innovative approaches? *(Health Check objective)*
11. To what extent has the RDP contributed to creation of access to broadband internet (including upgrading)? *(Health Check objective)*
12. To what extent has the NRN contributed to RDP objectives?
13. To what extent has the TA contributed to RDP objectives?
14. How efficiently have the resources allocated to the RDP been used in relation to achieving the intended outputs?

Measure-related CEQs

For each of the Axis 1 measures included in the RDP:

15. How and to what extent has the measure contributed to improving the competitiveness of the beneficiaries? *(where relevant, the answers to this CEQ should be presented so that the contribution to the competitiveness of the agriculture and forestry sectors can be seen separately)*

For each of the Axis 2 measures included in the RDP:

16. How and to what extent has the measure contributed to improving the environmental situation?

For each of the Axis 3 (Article 52(a)) measures included in the RDP:

17. How and to what extent has the measure contributed to the economic diversification of the beneficiaries?

For each of the Axis 3 (Article 52(b)) measures included in the RDP:

18. How and to what extent has the measure contributed to the improving the quality of life of beneficiaries?

For each of the Axis 3 (Article 52(c) and (d)) measures included in the RDP:

19. To what extent has the measure enhanced beneficiaries' capacities to improve economic diversification and quality of life in rural areas?

For each measure included in Axes 1-3 of the RDP

20. What other effects, including those related to other objectives/axes, are linked to the implementation of this measure (indirect, positive/negative effects on beneficiaries, non-beneficiaries, local level)?

Axis 4 (Leader)-related CEQs

21. To what extent has the RDP contributed to building local capacities for employment and diversification through LEADER? *(Community strategic priority)*
22. To what extent have LAGs contributed to achieving the objectives of the local strategy and the RDP?
23. To what extent has the Leader approach been implemented?
24. To what extent has the implementation of the Leader approach contributed to improving local governance? *(Community strategic priority)*

The axis-related result indicators should be used in the evaluation at two levels: as a contribution to answering the programme-level evaluation questions, in conjunction with the relevant impact indicators, particularly in relation to establishing the net impact of the RDP, and at axis level, where they can demonstrate the achievements of each axis, and contribute to assessing the relative contributions of the measures implemented under each axis.

3 OUTLINE OF THE *EX POST* EVALUATION REPORT[161]

1. Executive summary

- Main findings of the evaluation.
- Conclusions and recommendations.

2. Introduction

- Purpose of the report.
- Structure of the report.

3. The Evaluation Context

- Brief contextual information about the programme: related national policies, social and economic needs motivating assistance, identification of beneficiaries or other target groups.
- Description of the evaluation process: recapitulation of the terms of reference, purpose and scope of the evaluation.
- Brief outline of previous evaluations related to the programme.

4. Methodological Approach

- Explanation of the evaluation design and the methods used.
- Description of key terms of programme-specific and common evaluation questions, judgement criteria, target levels.
- Sources of data, techniques for data collection (questionnaires, interviews; size and selection criteria for samples, etc.); information about how the indicators are calculated in order to assess the quality and reliability of the data and identify possible biases.
- Techniques for replying to the evaluation questions and arriving at conclusions.
- Problems or limitations of the methodological approach.

5. Description of Programme, Measures, and Budget

- Programme implementation: actors involved, institutional context.
- Composition of the programme; description of priorities and measures.
- Intervention logic of single measures.
- Budget foreseen for the entire programming period.
- Uptake and budget actually spent.

6. Answers to Evaluation Questions

- Analysis and discussion of indicators with respect to judgement criteria and target levels referred to by evaluation questions.

[161] Chapter 7 of the CMEF Guidance note B – Evaluation Guidelines

- Analysis and discussion of quantitative and qualitative information from public statistics, specific surveys or enquiries, or other sources.
- Answers to the evaluation questions.

7. Conclusions and Recommendations

- Coherence between the measures applied and the objectives pursued; balance between the different measures within a programme.
- Degree of achieving programme-specific objectives as well as objectives set out in the national strategy and the Community strategy.
- Recommendations based on evaluation findings, including possible proposals for the adaptation of programmes.

4 EXAMPLE OF A QUALITY ASSESSMENT GRID FOR THE EVALUATION REPORT

Title of the evaluation:

Department/unit responsible:

Evaluator/contractor:

Assessment carried out by: (*name organisations/units involved in the assessment)*

Date of quality assessment:

1) RELEVANCE

- ***Does the evaluation respond to information needs of the commissioning body and fit the terms of reference?***

SCORING

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

2) SCOPE

- ***Is the rationale of the programme and its set of outputs, results and impacts fully examined, including both intended and unexpected policy interactions and consequences?***

SCORING

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

3) APPROPRIATE DESIGN

- ***Is the design for the evaluation adequate for obtaining results needed to answer the evaluation questions?***

SCORING

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

4) RELIABLE DATA

- ***Are primary and secondary data collected adequate for their intended use and have their reliability been ascertained? Have data weaknesses and limitations been explained?***

SCORING

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

5) SOUND ANALYSIS

- ***Are qualitative and quantitative data appropriately and systematically analysed to answer evaluation questions and cover other information needs in a valid manner? Are cause and effect links between the intervention and its results explained? Are external factors correctly taken into consideration?***

SCORING

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

6) CREDIBLE FINDINGS

- ***Do findings follow logically from and are justified by the data or information, analysis and interpretations based on pre-established criteria? Are findings based on carefully explained assumptions and rationale?***

SCORING

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

7) VALID CONCLUSIONS

- ***Are conclusions non-biased and fully based on findings? Are conclusions clear, clustered and prioritised?***

SCORING

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

8) HELPFUL RECOMMENDATIONS

- ***Are areas needing improvements identified in coherence with the conclusions? Are the suggested options realistic, impartial and sufficiently detailed to be operationally applicable?***

SCORING

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

9) CLARITY

- ***Is the report well structured, balanced and written in an understandable manner?***
- ***Is the report easy to read and has a short but comprehensive summary? Does the report contain graphs and tables?***

SCORING

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

OVERALL ASSESSMENT OF THE FINAL REPORT

Overall, the quality of the report is assessed to be:

Poor	Satisfactory	Good	Very Good	Excellent

- Does the evaluation fulfil the contractual conditions?
- Are the findings and conclusions of the report reliable, and are there any specific limitations to their validity and completeness?
- Is the information in the report potentially useful for communicating the impacts and achievements of the programme?

5 EXAMPLE OUTLINE OF THE TECHNICAL SPECIFICATION OF THE TERMS OF REFERENCE FOR THE *EX POST* EVALUATION[162]

Context	**Overall purpose of the contract** The object of procurement is the *ex post* evaluation of the Rural Development Programme (RDP) in [Member State or region/regions] 2007-2013 according to Council Regulation (EC) No 1698/2005 and Commission Regulation (EC) No 1974/2006 and national requirements (*and other specific tasks, such as thematic studies).* **Context** Background and context for the evaluation. **Reference to relevant legislation** All the relevant regulations, guidelines, directives and national legislation should be mentioned.
Scope of the *ex post* evaluation	**Programme** Name of the RDP. **Description of the RD measures** Measures used in the RDP. **Geographical area concerned** Member State/Region in question. **Programming period** Programming period 2007-2013. **Specific focus intended** Specific areas, approaches, sectors, etc.
Objectives of the evaluation	**Common objectives for the *ex post* evaluation** The *ex post* evaluation is carried out to take stock of the results, achievements, impacts, efficiency, effectiveness of the programme. **Programme-specific objectives for the *ex post* evaluation**
Tasks to be performed	**Specific tasks or content of the *ex post* evaluation** The *ex post* evaluation should appraise - All aspects covered by Article 86 of Council Regulation (EC) No 1698/2005 and Council Decision 2006/144/EC on Community strategic guidelines to rural development and Council Regulation (EC) 79/2009 - Degree of utilisation of resources; - Effectiveness of the programme; - Efficiency of the programme; - Socio-economic impact of the programme; - The programme's impact on the Community priorities;

[162] Adapted from Annex 1 of the Guidelines on the Mid-Term Evaluation of the Rural Development Programmes 2007-2013, European Evaluation Network for Rural Development, July 2009 at

	- Factors contributing to success or failure of the programme; - The Common Evaluation Questions (list). Furthermore, the following aspects - Specific provisions (e.g. Leader, NRN, TA); - Programme-specific evaluation questions (list). Specific tasks of the evaluator with regard to structuring, observing, analysing or judging.
Methods	In order to come to robust and reliable conclusions based on representative data, well-known and tested methods should be used for the *ex post* evaluation. The *ex post* report has to explain the used methods and data sources and their implications for the quality of the data and results. This should allow an assessment of the reliability of the findings of the *ex post* evaluation and facilitate the provision of usable and sound conclusions and recommendations. The ToR may specify some methods or preliminary consideration on methodology that should be used in the evaluation but in general the methodological choices are decided only after the tendering process. In the proposal, the applicant has to describe and explain the intended methodological approaches for the *ex post* evaluation. The evaluation is expected to utilise already existing data as far as possible. The Managing Authority will support the evaluator in retrieving relevant data from other institutions. The Managing Authority will accompany the realisation of the *ex post* evaluation and will wish to be kept informed about the status of the evaluation regularly [usually six weeks]. The contractor may be asked to participate in events and to give presentations.
Timing and content of deliverables	**Deliverables** [The deliverables have to be defined according to the specific timetable and tasks of the *ex post* evaluation]. Deliverables can include an inception report, progress report(s), draft final report, final report, meetings, presentations. **Requirements regarding content, style, format, language, and structure** The result of the *ex post* evaluation has to be presented in a final report bringing together all elements of the evaluation. The report must be clearly structured and formulated and include an executive summary. The proposed *ex post* evaluation report structure should feature in the Annex of the ToR. A proposed table of content for the *ex post* evaluation can be found in **Part III:** Item 3. **Timeline for all deliverables**
Organisation of work	**Budget** - Maximum overall budget or budget range. **Information on** - Which costs can be covered; - How the remuneration is planned along the defined deliverables and the timeline; - How the deliverables will be approved by the client; - If and how and on which basis supplementary works would be remunerated in terms of time (prolongation of the task) and in terms of justified complexity of applied methods in favour of high quality of the *ex post* evaluation; - Responsibilities for contract management; - Interaction with the Managing Authority and possible steering group.

Sources and documentation	**List of appropriate sources and materials** Programme documentation (Regulations 1698/2005 and 1974/2006 and national decrees, National Strategy Plan, Rural Development Programme, *Ex ante* Evaluation, Mid-term Evaluation, other evaluations, studies and assessments, CMEF, known data sources, etc.). **Provisions for data protection and data security**
Required capacities and content of the offer	This part of the tender defines the required capacities and the requested evidence and explanations to check expertise, reliability and capability of the tenderer. This might include: - Formal qualification of the experts involved; - Their qualification in the field of rural development and related EU policies and their evaluation; - Multiannual expertise and professional experience in the area of evaluation of Rural Development Programmes; - Proven results in evaluation of EU rural development policy or other policies (evaluation reports, studies, publications); - Specific skills (e.g. methods) or knowledge (e.g. local development, environment); - Good communication and collaboration skills (e.g. via proven participation in international, interregional or other partnership-based project). Moreover, the content of the offer is defined by reference to a) The methodological approach; b) The organisation of the work and time plan; c) The cost plan; d) Formal specifications.
Tendering procedures and contractual clauses	**Description of the tendering procedure** - Manner of submitting tender (e-mail, letter, fax) and address for submissions; - Deadline for submitting the tender; - List of eligibility, selection and awards criteria for choosing the evaluator and possible weightings of the award criteria; - Communication of successful and/or unsuccessful tenders. **Relevant contractual clauses**

6 RETRO PLAN IN THE *EX POST* EVALUATION

TASK	DURATION	DEADLINE
Report submission to the Commission		**31.12.2016**
National/regional approval procedures		
Quality assessment of the final report		
Final report submission to MA/SG		
Draft final report submission to MA/SG		
Interim report submission to MA/SG		
Inception report submission to MA/SG		
Signing of *ex post* evaluation contract		
Assessment of tenders		
End of call for tender		
Launch of call for tender		
Preparation of call for tender		
Terms of reference		
Evaluation mandate		
(if no permanent evaluation steering group: Set-up of an evaluation steering group)		
Screening of data and information sources		
Screening of evaluation questions and indicators		
Assessing evaluation needs		
Planning the process and schedule for the evaluation		

7 EXAMPLES OF LEADER-SPECIFIC ADDITIONAL INDICATORS

Evaluation need (subject?)	Evaluation question	Proposed judgment criteria	Common CMEF indicators	Proposed additional indicator (results)
Leader method (7 principles)	"To what extent has the Leader approach been implemented?"	Area-based local development strategies intended for well-identified sub-regional rural territories are developed and implemented.	OI: Total size of the LAG area. OI: Total population in LAG area.	Share of rural territory covered by LAGs. Share of rural population covered by LAGs.
		Local public-private partnerships are established.	OI: No of local action groups.	No and share (%) of various types of partners in partnerships from sector, territory, institutional socio- economic, gender point of view.
		The bottom-up approach is implemented with decision-making power for LAGs concerning the elaboration and implementation of local development strategies.		No of community based meetings, events or initiatives undertaken to prepare local strategy broken down by community, municipality and LAG level. No of people and their share of the adult rural population participating in the preparation of local development strategy broken down by community, municipality and LAG level and by sectors (public, private, civil), out of it those attending more than 50% of meetings. Level of decentralisation: No and types of tasks of the total No and types of tasks in the implementation of Leader left at LAG level[163]. Number of LAG members actively participating in decision-making. Division by sector (private, civil, community, public) and proportion of meetings attended <50%>50%.
		Multi-sector design and implementation of the strategy based on the interaction between actors and projects of different sectors of the local economy are realised.		No of specific thematic focuses of local strategies divided by types.
		Innovative approaches are implemented.		Share of innovative projects (innovation as defined by RDP or/and LAG) on the total No of projects implemented by LAG/LAG beneficiaries.
		Cooperation projects are prepared and implemented.	OI: No of cooperation projects; OI: No of cooperating	No of cooperation projects per LAG. No of cooperation projects divided by types.

[163] http://enrd.ec.europa.eu/leader/leader/en/leader-focus-group_en.cfm

Evaluation need (subject?)	Evaluation question	Proposed judgment criteria	Common CMEF indicators	Proposed additional indicator (results)
			LAGs.	Average number of partners by type (public, private, civil) per cooperation project.
		Local partnerships are well networked.		No of LAGs in NRN. Number of newly established external relationships to key rural development stakeholder organisations e.g. through involvement in or contribution to project or LAG activity. No of other networks in which LAGs take part. Share of LAGs participating in other networks (per network) on the total number of LAGs; Average No of networks in which LAGs are participating (No networks/LAGs).
Leader contribution to governance and local capacities.	To what extent has the implementation of the Leader approach contributed to improving local governance? *(Community strategic priority)*	Arrangements for setting up and managing Leader at national/regional level are established.	OI: Total size of the LAG area. OI: Total population in LAG area.	Share of rural territory covered by LAGs. Share of rural population covered by LAGs.
		Arrangements for monitoring and evaluation as part of Leader governance and transparent implementation at national/regional level are established.		No and focus of evaluations of Leader approach.
		Arrangements for setting up and managing Leader at local level are established.	OI: No of actions supported. OI: No of projects financed by LAGs. OI: No of beneficiaries supported.	Percentage of total applications submitted which are successful. Average time period needed from the application submission to the contract with beneficiary.
		LAG actively encourages the community involvement.		% of staff time spent on community animation
		Arrangements for monitoring and evaluation at local level are established and implemented.		No of LAGs which undertake formal self-assessment of the implementation of their local strategy.

Evaluation need (subject?)	Evaluation question	Proposed judgment criteria	Common CMEF indicators	Proposed additional indicator (results)
		Quality of local governance increased.		% of local contributions to the LAG budget. No and types of project which LAG is doing apart of Leader subsidy. *Also see indicators for local action group*
		LAG members and beneficiaries satisfaction in LAG activities has increased.	OI: No of beneficiaries supported.	No of LAG members.
		Local actors empowered.		No and types of events where local actors participate in LAG decisions at community, municipality and LAG level per LAG by various target groups (LAG members, non-members, beneficiaries, etc.). No of local actors involved in LAG decisions at community, municipality and LAG level per LAG by various target groups (LAG members, non-members, beneficiaries, non-beneficiaries, etc.).
		Coordination between different level of governance improved;		Average No of interactions between LAG, regional and MA/PA per LAG. No of feedbacks which MA/PA received from LAGs in consultations of Leader guidance documents and of it those accepted.
	To what extent has the RDP contributed to building local capacities for employment and diversification through LEADER? *(Community strategic priority)*	Capacities to generate sustainable employment obtained through Leader and established.	RI: Gross number of jobs created. II: Employment creation.	No of jobs created and sustained due to LAGs activities divided according economic sectors (agriculture, forestry, food processing, other processing, services and tourism).
		Capacities for business start up and sustainable development obtained through Leader and implemented.		No of new start-ups due to LAG activities divided according economic sectors (agriculture, forestry, food processing, other processing, services and tourism). No of expanded sustainable businesses due to LAG activities divided according economic sectors (agriculture, forestry, food processing, other processing, services and tourism).
		Capacities for development of sustainable diversified economic activities obtained through Leader and implemented		

8 EXAMPLES OF EVALUATION METHODS

8.1 Estimation of direct programme effects by means of ATT[164] indicator (*Ex post* evaluation RDP 2000-2006, Hessen, Germany)

Background

The study[165] was conducted for the *ex post* evaluation of the RDP (2000-2006) in Hessen (DE). The impact of the less favoured area (LFA) scheme and agro-environment (AE) scheme on on-farm employment (calculated in FTE[166] and FTE per 100 hectare farmland) was analysed at the level of individual farms.

Method

The analysis is based upon bookkeeping data (similar to FADN) of approximately 450 farms in Hessen. Bookkeeping data comprise information on farm, farmer and farm household characteristics as well as on the participation in the LFA and AE scheme (amount of LFA/AE payments received per farm/year). For each of the 450 farms, data from 2000-2005 is available. During that time 89 (107) farms participated in LFA (AE programs). *Beneficiaries of LFA* (LFA=1) are defined as those farms that have received a positive amount of LFA payments in all six years (2000-2005). *Beneficiaries of AE* (AE=1) are farms that have not received any AE payments in the base year (2000) and have received AE payments from at least 2002 continuously up to 2005. *Non-beneficiaries* (LFA=0, AE=0) are farms that did not receive LFA or AE payments from 2000-2005. Farm level data was supplemented by regional data at NUTS 3 level (e.g. unemployment rate, industrial wages). The table below shows the structure of the data set used.

Table 12 Data structure

Farm	AEP	LFA	Farm FTE_00	Farm FTE_05	Rate of change	Farm size_00	other variables
1	1	1	1.3	1.2	-0.1	70	…
2	0	0	2.5	2.5	0	50	…
3	1	0	0.7	0.7	0	20	…
4	0	1	1.7	1.9	0.2	130	…
…	…	…	…	…	…	…	…

Note: Hypothetical figures

Employment effects of LFA and AE schemes are analysed by applying a control group comparison (DID) combined with Propensity Score Matching (PSM). A greedy matching estimator employing 1:1 matching without replacement was used to assign beneficiaries and non-beneficiaries to each other (Parson, 2001).[167]

In a first step, the propensity score is estimated using logistic regression. The dependent variable is the participation status of each farm. Explanatory variables included farm specific characteristics (e.g. farm size, share of grassland, expenditures for pesticides, livestock densities, farm profit, farm sales, age and education of the farm operator) and regional characteristics (unemployment rate, real estate prices, land rents, industry wages, etc.) in the base year 2000. The propensity score is the conditional probability of each farm to be a beneficiary and takes on values between 0 and 1.

[164] Average Treatment effect on the Treated
[165] Pufahl and Weiss, 2009; Farm Structure and the Effects of Agro-Environmental Programs: Results from a Matching Analysis for European Countries, University of Kent, Cantenbury, UK, 2009
[166] FTE – full time job equivalent
[167] Parson, L. S. (2001): Reducing bias in a propensity score matched-pair sample using greedy matching techniques. Proceedings of the 26th Meeting of the SAS International User Group (SUGI 26): www2.sas.com/proceedings/sugi26/p214-26.pdf.

The selection of relevant explanatory variables was guided by prior information about what factors drive participation to LFA and or AE schemes. For instance, AE beneficiaries (on average) manage their land (prior to participation) less intensively than non-beneficiaries. Thus, indicators proxying land use intensity (e.g. livestock density, share of extensive crops/grassland, expenditures for pesticides and fertiliser) are considered. Participation to LFA depends on whether a farm is situated inside or outside a disadvantaged area. Thus, criteria used to delineate the eligible area (e.g. soil and climate conditions, topography) supplemented by farm characteristics replicate factors influencing the decision to participate. Furthermore, using FTE (FTE per 100 hectare) as an explanatory variable ensures that both groups have a similar level of on-farm employment in the base year.

Beneficiaries and non-beneficiaries with similar propensity scores are assigned to each other. Only those 68 beneficiaries for which non-beneficiaries with similar propensity scores are available were used for further analysis. 39 beneficiaries had to be excluded from the analysis because no similar non-beneficiaries were available (because the pool of non-beneficiaries was too small). This violates the common support assumption[168] and may lead to biased results. After matching, it needs to be checked whether farm characteristics of beneficiaries and non-beneficiaries are on average similar (See table below). Differences in means were tested using a T-Test.

Table 13 Means of selected characteristics before and after matching (AE programme)

Characteristic in base year (2000)	Unit	Before matching		After matching	
		Beneficiaries	Non-beneficiaries	Beneficiaries	Non-beneficiaries
Farmland	hectare	**71.29**	**45.67 ***	59.01	50.80
On-farm labour	FTE	**1.49**	**1.32 ***	1.44	1.39
On-farm labour/100 hectares	FTE	**2.09**	**2.89 ***	2.44	2.28
Farm income	EUR 1,000	**22.91**	**16.02 ***	18.89	16.15
Number of observations		**107**	**237**	68	68

Note: Asterisks indicate significant differences in means between beneficiaries and non-beneficiaries in the base year 2000 at < 5% level as measured by the T-Test.

The last step involves the calculation of the rate of change of on-farm employment between 2000 and 2005. Mean rates of change are computed for beneficiaries and non-beneficiaries (See columns (3) and (4) in the table below). The Average Treatment effect on the Treated (ATT) is the difference between the rates of change of both groups ((1)-(2)). The T-Test is used to detect whether observed differences in means are significant.

Table 14 Employment impact of the LFA and the AE scheme in Hessen, Germany (2000-2005)

Indicator of interest	Unit	Rate of change Beneficiaries	Rate of change Non-beneficiaries	Average Treatment effect on the Treated (ATT)	
		(1)	(2)	(1)-(2)	
Less favoured area scheme		(n=54)	(n=54)		
On-farm labour	FTE	-0.03	-0.14	0.11	*
On-farm labour/100 hectares	FTE	-0.36	-0.15	-0.21	
Agro-environment scheme		(n=68)	(n=68)		
On-farm labour	FTE	-0.16	0.00	-0.17	***
On-farm labour/100 hectares	FTE	-0.42	0.02	-0.44	***

Note: *** (**,*): significant at the 1% (5%, 10%) level.

[168]The assumption of common support states, that there should be a similar non-beneficiary for each beneficiary.

Results

Beneficiaries of LFA show almost no change in total on-farm employment from 2000-2005 while non-beneficiaries reduced total on-farm employment by 0.14 FTE per farm on average. The difference between both groups results in an average effect for beneficiaries (ATT) of +0.11 FTE and is significant at a 10 per cent level. Against the general decrease in agricultural employment, LFA helped to maintain labour in agriculture. The opposite is true for AE programs, because support AE measures in Hessen predominantly support (labour) extensive production methods. This results in an average reduction of agricultural labour force of 0.17 FTE per farm (significant at 1% level). On farm labour per hectare also decreases strongly under AE participation (-0.44 FTE per farm, significant at 1% level) because the amount of farmed land is extended significantly under AE programs[169].

8.2 Estimation of direct programme effects occurring at the level of SAPARD programme beneficiaries (Slovakia)

Background

PSM methodology was applied[170] in order to assess the results of the RD SAPARD programme in Slovakia (Measure 1: Support of investment in agricultural enterprises) on programme beneficiaries. The programme support under Measure 1 was primarily targeting the following agricultural sectors: a) beef sector, b) pig sector, c) sheep sector, d) poultry sector, e) fruits and vegetables sector. Programme support under Measure 1 had the form of a capital grant covering up to 50% of the costs to investments in the above sectors. The major beneficiaries of programme support (they received approximately 67% of funds available under this measure) were large agricultural companies located in relatively well developed regions of West Slovakia. The assessment of programme impacts on the agricultural companies was carried out on the basis of Slovak FADN database in the years 2002-2005.

Method

The following methodological steps were carried out in order to estimate the direct programmes effects on beneficiaries:

(a) SAPARD beneficiaries were identified and selected from the existing FADN databases to the panel. Data for each SAPARD beneficiary was collected in the years 2002-2003 (i.e. prior to their participation in SAPARD) and 2005 (after implementation of SAPARD).

(b) SAPARD general and specific eligibility criteria (e.g. pre-defined farm performance coefficients and farm profitability ratios; various minimum/maximum production, age, etc. thresholds, etc.) that were valid in individual years were translated into respective quantitative coefficients and applied to all non-SAPARD units included in FADN databases.

(c) Units which satisfied the above criteria in the years 2002-2005 and did not receive support from SAPARD programme were selected to the panel of a control group (*eligible* non-participants).

(d) Respective balanced panels (i.e. embracing SAPARD beneficiaries and all non-SAPARD units meeting SAPARD eligibility criteria in specific years) were constructed for the years 2002-2005, i.e. observations on the same units in period 2002-2005.

On the basis of available Slovak FADN database, 232 agricultural companies were selected for further analysis (balanced panel data) which was performed for the years 2003 (before SAPARD) and 2005

[169]Pufahl and Weiss, 2009; Farm Structure and the Effects of Agro-Environmental Programs: Results from a Matching Analysis for European Countries, University of Kent, Cantenbury, UK, 2009
[170]Michalek J. (2012a), "Counterfactual impact evaluation of EU Rural Development Programmes - Propensity Score Matching methodology applied to selected EU Member States", *Volume 1 – A micro-level approach*.", European Commission, JRC Scientific and Policy Reports, pp 1-95, http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=5379

(after SAPARD). Out of the selected 232 agricultural enterprises, there were 51 agricultural farms SAPARD participants and 181 SAPARD non-participants but yet, SAPARD eligible.

The preliminary analysis showed that agricultural companies that received support from the SAPARD programme differed significantly in a number of important characteristics from eligible programme non-participants, i.e. SAPARD beneficiaries were in general much larger (ha), they employed more people and were more profitable (i.e. less unprofitable) compared to the agricultural companies that were non-supported from SAPARD. Given above, the group of SAPARD participants could not be directly compared to non-participants, as a selection bias would have been incurred.

In order to ensure comparability, the PSM method was applied using number of individual characteristics of agricultural companies as covariates (methodology applied to the selection of variables into the estimated logit function is described in the cited study). Imposition of common support region and selection of appropriate matching algorithm resulted in dropping from a further analysis the companies that were non-comparable. The applied balancing property tests (t-test) showed that the selected matching procedure (i.e. kernel epanechnikov bandwidth 0.06) considerably improved comparability of both groups of agricultural companies, making a counterfactual analysis more realistic. Indeed, there were previously significant differences in the most important farm characteristics between the group of agricultural companies supported from the SAPARD programme (D=1) and non-supported farms (D=0) dropped after matching (differences became no more statistically significant). This applies to all important variables determining both programme participation and outcomes, e.g. profit per company (prior to SAPARD programme), liabilities, value of current assets, etc.

The assessment of the micro-economic effects of a given RD programme on programme beneficiaries was carried out in both groups of farms using seven results indicators available from a standard FADN system:

- Profit per company,
- Profit per ha,
- Profit per person employed,
- Gross value added (GVA) per company,
- Employment per company,
- Labour productivity (Gross value added per employed),
- Land productivity (Gross value added per ha).

The results also showed that simple techniques applied to the estimation of programme effects can be highly misleading, whereas the application of advanced evaluation methodologies can lead to quite different yet much more reliable results.

This issue can be very well illustrated on the basis of the table below.

Table 15 Estimation of the direct effect of RDP on agricultural companies (comparison of various methods)

Methods/units	GVA/company In SKK[171] 1,000		
	Before programme (T_0)	After programme (T_1)	DID (difference between T_1 and T_0)
Programme participants (P=1)	17,727	18,478	751
Programme non-participants (P=0)	9,950	9,680	-270
Average Ø	11,660	11,614	-46
Difference (1-0)	7,777	8,798	1,021
Difference (1- Ø)	6,067	6,864	797
Matched programme participants (M=1)	11,082	9,610	-1,472
Matched programme non-participants (M=0)	9,367	9,701	334
Average Treatment Effect on Treated (ATT) using Propensity Score Matching (PSM)	1,715	-90	-1,805

Where:

(P=1) A group consisting of all direct programme beneficiaries (before matching);

(P=0) A group consisting of all programme non-beneficiaries (before matching).

(1-0) Programme effects calculated as a difference in outcome indicator (here: GVA/company) between direct programme beneficiaries (P=1) *and non-beneficiaries* (P=0).

(1- Ø) Programme effects calculated as a difference in outcome indicator (here: GVA/company) between direct programme beneficiaries (P=1) *and an average Ø (country or sample).*

(M=1) A matched group of direct programme beneficiaries.

(M=0) A matched group of programme non-beneficiaries.

(ATT) Average Treatment on Treated calculated as a difference between (M=1) and (M=0).

DID Difference in differences between outcomes observed in periods T_1 and T_0.

Results

Analysing the figures in the table above shows that:

- If a naïve before/after estimator was applied, the effect of the programme would be assessed as very positive (average change in GVA per company = +751 thousand SKK). Yet, this estimator is statistically biased.
- If programme participants were compared with all (unmatched) programme non-participants (before and after) and DID estimator was applied, the effect of the programme would also be assessed as very positive (average change in GVA per company = +1,021 thousand SKK). Yet, this estimator is statistically biased.
- If effects observed for programme participants were compared with a country's average (e.g. performance standards) calculated for all farms, i.e. programme participants and non-participants (before and after) and DID estimator was applied the effect of the programme would be assessed as very positive (average change in GVA per company = +797 thousand SKK). Yet, similar as in (1) and (2) this estimator is statistically biased.
- The conclusions above have to be revised in case the programme effects are measured using statistically similar (matched) groups (participants vs. non-participants). In this case the estimated programme effect (DID in ATT) was found to be negative (!) (average change in GVA per company = -1,805 thousand SKK). The reason is a much higher growth in GVA per

[171] SKK = Slovak crowns

company in the similar (matched) group of SAPARD non-participants (average change in GVA = +334 thousand SKK) compared with the matched SAPARD participants (average change in GVA = -1,472 thousand SKK).

The above example clearly shows that application of an incorrect (naïve) methodology to assess programme effects (e.g. a micro-economic effect of an investment support measure) may lead to a significant bias in estimated programme results making them unreliable.

8.3 Estimation of leverage effects for RDP farm investment measure (Schleswig-Holstein, Germany)

Background

A leverage effect can be considered as an important micro-economic consequence of a RD support. It means the extent of additional private or business-related spending which is induced by public funding (e.g. in form of RD programme).

The choice of an appropriate indicator depends on the focus of the evaluation: if the focus is on increasing productivity of the agricultural sector, *the amount of money transferred for other (i.e. non-supported) investments* will be a good proxy; if the focus is on the regional economy in general, private consumption can be included as well.

Method

The leverage effect can be calculated by taking the following methodological steps[172]:

(a) selection of individual units j supported by a RDP;

(b) identification of a comparison/control group m matching with units j (identical distribution of covariates) in the period T=0 (i.e. prior to j access to the programme) using PSM method;

(c) selection of relevant result indicators as proxies for private or business-related spending, e.g. *money transfers from farm to farm household*; *level of private and farm consumption*, etc.;

(d) calculating ATT for selected result indicators between both groups (i.e. j and m);

(e) Applying DID on the estimated ATT.

It is expected that in case of a significant leverage effects the calculated DID-ATT will be positive and significant.

The following example illustrates how the above methodology was applied to estimate the leverage effects in the RD Agrarinvestitionsprogramm in Schleswig-Holstein (Measure: Investments in milk and beef sectors). In this particular case, the leverage effects are calculated on the regional economy, using the *money transfers for private consumption and building of private assets* as indicators.

Results

The results obtained on the basis of 1,333 bookkeeping farms (101 AFP[173] participants and 1,232 non-participants) specialised in milk production (panel for years 2001-2007) indicates significant leverage effects (Michalek, 2012), i.e. participation in AFP programme resulted in significant additional transfers of funds from farms to household (in average EUR 4,653 per farm; or EUR 3,178 per farm), thus induced private spending (See below).

[172]Michalek J. (2012a), "Counterfactual impact evaluation of EU Rural Development Programmes - Propensity Score Matching methodology applied to selected EU Member States", *Volume 1 – A micro-level approach*.", European Commission, JRC Scientific and Policy Reports, pp 1-95, http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=5379
[173]AFP - Agrarforderungsprogramme

Table 16 **Estimation of the leverage effects in AFP programme (Schleswig-Holstein): *Result indicator- Money transfer from farm to farm household for living***

Calculation basis	Variable: Money transfer from farm to farm households for living		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	30,072	43,810	13,738
Unmatched P=0 (1,232)	24,512	32,336	7,824
Ø (1,333)	24,933	33,206	8,273
Difference (1 minus 0)	5,560	11,473	5,913
Difference (1- average Ø)	5,139	10,604	5,465
Matched M= 1 (101)	30,072	43,810	13,738
Matched M= 0 (1,067)	27,647	36,732	9,085
ATT	2,424	7,077	4,653

Table 17 **Estimation of the leverage effects in AFP programme (Schleswig-Holstein): Result indicator- *Money transfer from farm for building of private assets***

Calculation basis	Money transfers from farm for building of private assets		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	18,447	48,302	29,855
Unmatched P=0 (1,232)	11,632	31,926	20,294
Ø (1,333)	12,148	33,167	21,019
Difference (1 minus 0)	6,814	16,376	9,562
Difference (1 minus average Ø)	6,299	15,135	8,836
Matched M= 1 (101)	18,447	48,302	29,855
Matched M= 0 (1,067)	17,504	44,181	26,677
ATT	942	4,120	3,178

8.4 **Estimation of multiplier effects using input-output model** (Castilia y Leon, Spain, RDP 2007-2013)

Background

The input-output method offers one alternative for assessing the impact of the 2007-2013 RDP at regional economy level, and more specifically on each of the productive sectors. It also provides an approach to its consequences for production, employment and added value. This method was applied in the assessment of the Integrated Operative Programme (2000-2006) and in the *ex ante* assessment of the Castilla y León (Spain) 2007-2013 RDP.

Method

The method used applies the Leontief open production model to determine the total effects (both direct and indirect) on regional sectoral added value generated by the increase in final regional sectoral demand due to the intervention. In short, it calculates

$$g = \hat{V}(I - A_d)^{-1} f_d^*$$

where:

$\boldsymbol{g}$ being the vector of the sectoral added value generated by the increases in final demand as a result of the intervention.

$\hat{v}$ being the diagonal matrix expressing added value absorbed by the sectors to produce one unit of product.

$\boldsymbol{A_d}$ is the matrix of internal technical coefficients.

$\boldsymbol{f_d^*}$ is the interior sectoral demand vector due to the intervention (net imports).

Finally, the added value obtained, expressed in PPS, is compared with the baseline situation (corresponding to the date of the input-output table) calculating the change in the added value as a result of the intervention.

In order to apply this method, the increase in final regional sectoral demand (f^*) originating from the intervention needs to be determined. This information is not directly observable and must be obtained from managers and coordinators of the RDP or from direct beneficiaries of each measure. For the case of Castilla y León we used both procedures and requested expenditure distribution (in percentage terms) amongst the 51 areas of activity into which the sectors in the *Castilla y León input-output table* for 2000 are grouped.

As interest is focused on ascertaining the impact of European support on the regional economy (Community added value), we needed to calculate the part corresponding to the increase in demand devoted to purchasing goods and services produced only in the region, in other words, the increase in interior sectorial demand (fd^*). To do so, we could also use the information provided by the *input-output table,* subtracting the part in the increase in demand of each sector to be met by imports (from other regions in Spain or other countries).

Furthermore, in order to make the effects of the various years comparable, we work with constant Euros (preferably from the year of the *input-output table*). We then deflate the expenditure of the different axes of the RDP over the implied years.

Advantages

(a) The method allows calculation of the direct and indirect effects for the measures both in aggregate terms as well as in sectors.

(b) The model may be applied for each axis, for each year and for the entire RDP over the whole period (2007-2013). Therefore overall effects throughout the whole period as well as the differences between the respective axes and years may be analysed.

(c) The method provides prior estimations of the indicator based on forecasts in expenditure distribution, thereby offering advantages from the point of monitoring the programme.

Drawbacks

(d) The method does not reflect possible changes in regional productive structure as it always uses the same input-output table. This proves to be even more of a drawback if the available input-output table offers data corresponding to a year some way back in time.

(e) Quantitative and qualitative techniques need to be merged to determine expenditure distribution by areas of activity. The information required must take account of the various agents in charge of executing the RDP (regional government, central administration, local councils, local action groups), meaning that it proves quite difficult to obtain and that the final results may be conditioned by the reliability of the responses provided.

(f) The effects determined in this approach are only domestic and thus restricted to the region. Moreover, the method assumes that the effects are produced in a year or that the effect is accumulated.

(g) The method does not take into account the leverage effects and deadweight of the action, although the procedure may be improved by estimating such effects through surveys and case studies.

Due to the heterogeneity of the programme itself, which covers a wide range of actions, the findings are restricted. In addition, certain measures are not aimed at boosting any particular sector of the economy but seek rather to impact on the whole economy, as a result of which their quantification by sectors proves difficult.

8.5 Assessment of Axis 1 impacts on the competitiveness of agriculture sector (Slovakia RDP 2007-2013)

Background

The assessment of the RDP Axis 1 - impacts[174] on competitiveness of the agriculture enabled to understand to what extent the RDP interventions under Axis 1 affected the situation in the competitiveness of supported farms and in the agriculture sector of the Slovak Republic within three years after the programme´s start. The assessment has considered the following indicators:

- CMEF: economic growth (net GVA), labour productivity and employment,
- RDP-specific: farm property, commitments, production, profit per farm/AWU.

The collected evidence and related judgments were used in answering CMEF horizontal evaluation questions No 1, 5, 7 and 9. Results of the assessment served as the important information source for the programme authorities, policy makers, RDP Monitoring Committee and other rural development stakeholders in enhancing the quality of the programme design and implementation.

Method

The combination of PSM[175], DiD[176] and ATT [177] has been employed in conducting counterfactual analysis used in the assessment.

The extensive data base built by RIAFE on the information sheets of the Ministry of Agriculture of Slovakia has enabled to apply the described method (PSM, DiD and ATT). The database was used as the data source to conduct the counterfactual analysis. This database has covered 82,7% of the LPIS SK within the period of 2006-2009, out of which 74,4% were legal persons (cooperatives, limited and share holding companies) and 8,3% physical persons (family farms). The reasons for selecting this database for the assessment of the RDP Axis 1 - impacts were as follows:

- Database covers 4 times bigger number of farms (2260 – 2613 farms sample) as FADN (600 farms sample),
- The size of the database allows to compare treated and no-treated farms in the timeline (FADN has too little sample for this purpose),

[174]The presented case was conducted within the mid-term evaluation of the RDP of Slovakia, conducted by the Research Institute of Agriculture and Food Economic (RIAFE) in Bratislava in collaboration with the private sub-contractor EuroConsulting ltd.
[175]Page 29, Approaches for assessing impacts of the Rural Development Programmes in the context of multiple intervening factors, European Evaluation Network for Rural Development, 2010, Brussels
[176]Page 30, Approaches for assessing impacts of the Rural Development Programmes in the context of multiple intervening factors, European Evaluation Network for Rural Development, 2010, Brussels
[177]Page 66, Approaches for assessing impacts of the Rural Development Programmes in the context of multiple intervening factors, European Evaluation Network for Rural Development, 2010, Brussels

- Database covers the whole territory of Slovakia and takes in consideration significant portion of agriculture land,
- Except of economic indicators the database considers the entire structure of support instruments of the EU Common Agriculture Policy and Slovak Aid for agriculture and rural areas (the same structure as FADN).

The structure and size of the RIAFE database used in the assessment of the RDP Axis 1 - impact in 2006 – 2009 is shown in Table 13:

Table 18 The structure of the RIAFE database, based on the information sheets of the Ministry of Agriculture

Farms	Numbers				%			
	2006	2007	2008	2009	2006	2007	2008	2009
Legal persons	1392	1374	1331	1421	59,5	59,0	58,9	54,4
Physical persons	949	956	929	1192	40,5	41,0	41,1	45,6
Total	2341	2330	2260	2613	100	100	100	100

Source: MTE of the RDP Slovakia, RIAFE, Euroconsulting, 2010

For the purpose of the assessment of the RDP Axis 1 - impacts the evaluator has created a sample of farms based on the above database. The composition of the sample was identical across the period of 2006 – 2009.

The structure of the sample is shown in Table 14

Table 19 Structure of the sample selected for the assessment of impacts

Farms	Numbers	%
Legal persons	1047	66,6
Physical persons	525	33,4
Total	1572	100

Source: MTE of the RDP Slovakia, RIAFE, Euroconsulting, 2010

To conduct the counterfactual analysis the sample was used to construct the control groups of treated and not-treated farms. The structure of control groups is shown in Table 15:

Table 20 Structure of counterfactual analysis

Legal form	Number			%		
	Legal person	Physical person	Total	Legal person	Physical person	Total
Treated	279	110	389	26,6	21,0	24,7
Non-treated	968	415	1183	73,4	79,0	75,3
Total	1047	525	1572	100	100	100

Source: MTE of the RDP Slovakia, RIAFE, Euroconsulting, 2010

“Treated” are farms which have received the support from investment measures of Axis1, RDP 2007 – 2013. “Non-treated” are those that have not received the support from Axis 1, however they could

receive support from previous programmes. The assessment period was 2006 - 2009, with baseline data collected in 2006. In addition each year of the period was analysed separately.

The assessment was conducted at following levels:

- NUTS I (Slovakia)
  - All farms
  - Legal persons
  - Physical persons
- NUTS III (8 regions)
  - All farms

The assessment was conducted as follows:

- Selection of impact indicators (based on the RIAFE database) using factor analysis,
- Estimation of prosperity matching score using impact indicators,
- Selection of matching method algorithm,
- Analysis of results for treated and non-treated farms,
- Calculation of DiD and ATT.

Based on factor analysis **the following indicators** have been selected for the assessment:

- GVA (Slovak crowns)
- Labour productivity (GVA/FTE)
- Employment (AWU),
- Farm property (Slovak crowns),
- Commitments per farm (Slovak crowns),
- Production per farm (Slovak crowns)
- Profit per farm, AWU.

The following indicators have been used in estimation of propensity matching score (PSM) and creation of the matching sample of treated and not-treated farms:

- Employment (AWU),
- Commitments (Slovak crowns),
- LPIS (Agriculture land registered in LPIS in ha),
- Production per farm (Slovak crowns),
- GVA (Slovak crowns).

As matching method algorithm was selected for Data Matching: Greedy software NCSS 2007, which is the method of searching the conformity of words in the chain “from one to another”. All above indicators and the indicators: profit per farm and labour productivity was used in this algorithm.

Based on this estimation the share of matching farms on the total sample used in the assessment was as shown in Table 16.

Table 21 Number and share of treated and non-treated farms in the period of 2006 – 2009

	Group	Number	matched	%matched	unmatched	% unmatched
2006	P1	389	311	79,95	78	20.05
	P0	1184	991	83,70	193	16,30
2007	P1	389	307	72,92	82	27,08

	Group	Number	matched	%matched	unmatched	% unmatched
	P0	1184	967	81,67	217	18,33
2008	P1	389	315	80,98	74	19,02
	P0	1184	998	84,29	186	15,71
2009	P1	389	308	79,18	81	20,82
	P0	1184	1001	84,54	183	15,46

Source: MTE of the RDP Slovakia, RIAFE, Euroconsulting, 2010

Results

Assessing the RDP Axis 1 effects on competitiveness of the agriculture sector in Slovakia has enabled to produce the following evaluation results:

- The property of supported farms has increased significantly compare to their non-supported pairs. In case of physical person, the increase of property would happen even without the aid.
- RDP beneficiaries had higher in-depth situation at the beginning of the programming period compare to others, however the difference has decreased at the time of the assessment, less in the case of physical persons compared to legal entities.
- The support was accompanied with a high deadweight rate (227%) and differentiation between treated and non-treated farms, which might have a consequence on the ability to absorb funds in the future.
- In relation to production, the treated farms have shown more production stability (mainly in case of physical persons) than their non- treated counterparts.
- In general the GVA has decreased in the agriculture sector, however it was less significant in the group of supported farms.
- The RDP support had negative influence on the farm profit (more in the case of legal persons and the physical persons), probably due to depreciation of newly purchased machinery. A similar situation was observed in the profit per AWU.
- The employment has decreased in both groups – treated and non-treated, however the decrease for supported farms, was lower than for those that have not received the aid.

8.6 Model-based assessment on impacts of the RDP on growth and employment (Austria)

Background

This good practice example takes a look at one of these studies[178] elaborated in the course of the mid-term evaluation of the Austrian RDP 2007-13[179] which has been submitted to the EC in 2010. The study was supposed to help understand if there is a causal relationship between the interventions

[178] Sinabell e.a. 2011
[179] As for Austria's Rural Development Programme, monitoring and evaluation tasks are housed in the Managing Authority, the Ministry of Agriculture, Forestry, Environment and Water.

funded by the RDP and the specific and relatively advantageous situation: slower declining number of farms compared to the EU average and higher than cities economic growth in rural areas.

The study did not only examine the possible effects of the overall programme on growth, employment and selected environmental media, it also helped create a sound database for further evaluations, notably the *ex post* evaluation but also beyond that horizon. The study is complemented by two corollary background papers; one to consolidate the indicator set for gender-based assessment of impacts; the other one to complete the database required for a total account for the agricultural sector at NUTS 3 level.

Method

The study adopts a two-pronged approach: (i) assessing the territorial effects of RDP measures, and (ii) assessing the RDP measures' effects on the farming sector and the wider economy. To carry out the first type of assessment, the evaluator chose a multi-regional econometric input-output-model called MultiReg which has first been applied to assess the impacts of EFRD interventions in Austria between 1995 and 2007[180]. As for the sectorial analysis, the evaluator chose the PASMA model (Positive agricultural and Forestry Sector model), that is frequently used in agricultural and environmental policy analyses[181]. Both models were then combined and complemented by a statistical analysis of the economic accounts of agriculture at NUTS 3 level[182] made specifically for this task.

The following key questions were guiding the study:

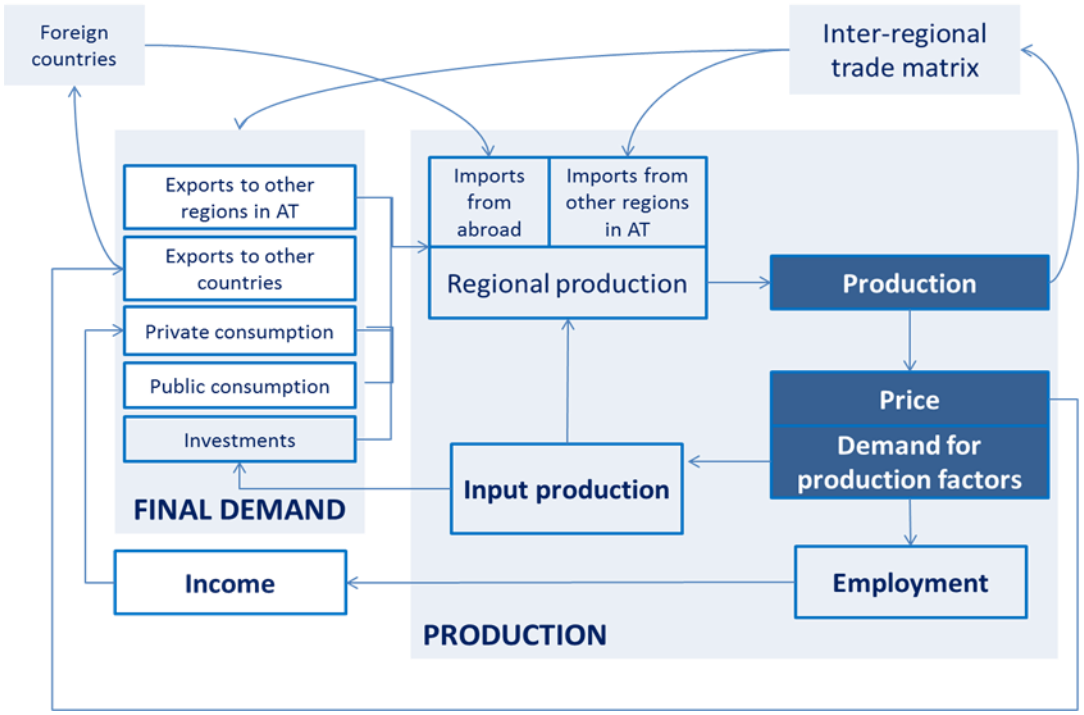
- To what extent does the RDP contribute to maintaining agriculture in disadvantaged areas?
- To what extent does the RDP improve the situation concerning agricultural incomes?
- To what extent does the RDP contribute to the extensification of agriculture and by virtue of this to the reduction of potential pressure on the environment?
- To what extent does the RDP promote the whole economy by increased provision of services into and an increased flow of products from the agricultural sector, although payments are mainly directed towards agriculture?
- To what extent does the RDP contribute to growth and employment not only in rural areas, but also in other (integrated and urban) areas?

Referring to the seven common impact indicators of the CMEF, the study directly addresses the three socio-economic ones (economic growth, employment and labour productivity); concerning the environmental indicators, it addresses the indicator on water quality and, in parts, the indicator on climate effects. All results have been calculated at NUTS 3 level. All in all, the study (fully or partly) addresses eight of the 19 Horizontal Evaluation Questions.

As yet the MultiReg model (See the presumed inter-relationships underlying this model in the diagram below) which captures the effects of economic changes in one region on those in other regions has been applied at Länder (NUTS 2) level. The availability of regionalised IACS[183] data allowed for breaking the model down to NUTS 3 level.

[180] Mayerhofer e.a. 2009
[181] Schmid et al. 2007; Schmid and Sinabell, 2007
[182] Statistik Austria, 2009
[183] called InVeKoS in Austria

Figure 19 The structure of the MultiReg model



Source: Joanneum Research and WIFO (from Sinabell e.a.2011)

Overcoming the limitations of quantitative statistical analysis, the architecture of the MultiReg model helps materialise the results chain and understand the cause-effect relationships. These relationships are of hypothetic nature; however they build, as far as possible, on empirical sources of knowledge, and only, where those are missing, on educated guesses. The model is inspired by a "Keynesian" perception of state interventions; it distinguishes between:

- Direct effects: changes in growth and employment of subsidised farms or firms;
- Backward linkages:
  - Indirect effects: changes through increased demand for supply and services;
  - Induced effects: changes in income and consumption expenses of people and households related to the beneficiary;
- Forward linkages (also called "catalytic effects"): changes through increased and productivity and competitiveness of subsidised farms or firms.

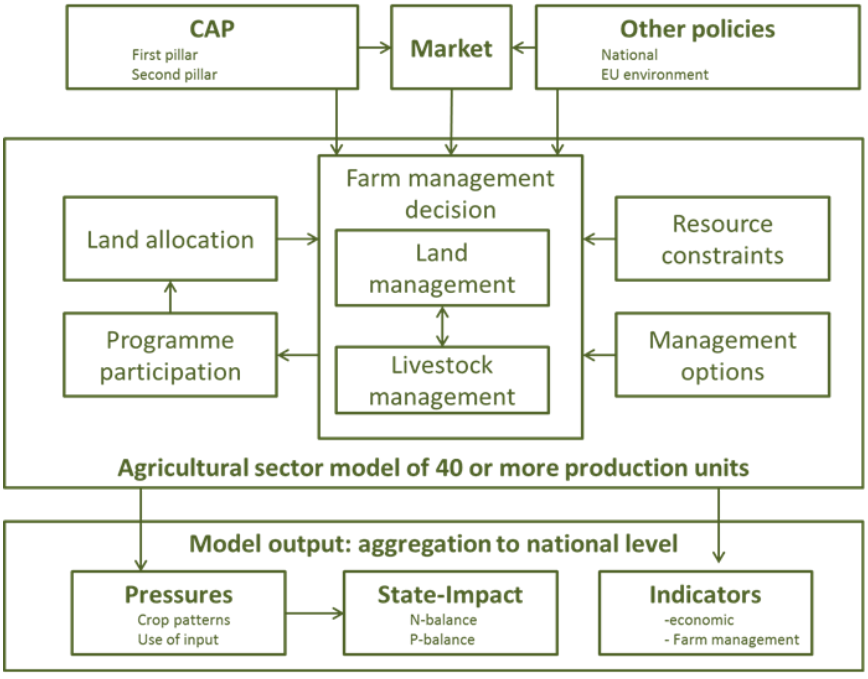
MultiReg had been developed by WIFO (Austrian Institute of Economic Research) together with the Joanneum-Research (at the Centre for Economic and Innovation Research) for 32 sectors/product categories and the nine Austrian Länder, modelling inter-linkages within Austria and with foreign countries.

For this evaluation study, the model has been further differentiated into 58 sectors/product categories and 99 political districts of Austria. MultiReg integrates regional input-output tables with a survey-based inter-regional trade flow matrix and with the analysis of time series which provided empirically based co-efficients interlinking significant variables (such as household income and demand, production volume and employment, etc.). Finally a model of the Austrian fiscal equalisation scheme (revenue sharing between regions and municipalities) and an estimation of social security payments have been added.

The non-linear programming PASMA model depicts the inter-linkages and flows of the Austrian agricultural system. It has been created as a policy advisory tool depicting the whole agricultural and forestry sector in terms of structure, production and policies. It is based on the methodology of Positive Mathematical Programming[184] using non-linear objective functions. The input data at NUTS 3 level originate from the IACS and various agricultural surveys. The advantage of this model is that it produces valid results even with smaller datasets than other models building on the aggregation of individual farm data. The following data has been fed in:

- Farm Structure Survey 1999 (at farm level);
- IACS data at farm and municipal level (time series between 2000 and 2008);
- Data on standard gross margins from 2002/03;
- Average required work time (labour input) based on surveys from 2002 and 2006;
- Statistical data, price information, etc. from various FAO, OECD, EU, Austrian and sector-specific sources;
- Specific parameters and co-efficients (e.g. for organic agriculture, from relevant research publications);
- Information from experts (e.g. from the advisory committee for the study).

Figure 20 The structure of the PASMA model



Source: WIFO (from Sinabell e.al 2011)

The combination of the two models allows for modelling sectorial and regional inter-linkages and flows, within and between regions, between rural and urban regions, between Austria and foreign countries; thus it allows for making conclusions on the impact of the RDP on the national economic system. Both models have been adapted to NUTS 3 and integrated in a way that the indicators could be calculated consistently and conclusively in accordance with the national accounting system and statistics.

The link between the two models has been framed with a set of plausible assumptions:

[184] Howitt 1995

- The period of observation is 2007-2013. The results only apply to this period;
- Prices are determined at international level; there is no influence of Austrian demand on price levels;
- Decreased production in the country is compensated by increased imports to maintain the same throughput in the processing industries;
- Decreasing demand from agriculture for supply hits these sectors as they mainly produce for the domestic market (e.g. construction);
- Productivity does not increase within this period; the assumption goes that the decrease of productivity during the investment phase and the increased productivity after start-up are balancing each other.

The effects of the individual RDP measures are either modelled by MultiReg (typical case: training, farm advisory or investment measures) or PASMA (typical case: LFA or agri-environmental measures). The results of PASMA flow into MultiReg, never the other way round.

To net out the effects of the RDP, two scenarios have been developed:

- Forecast on the developments in the agricultural and other economic sectors WITHOUT the RDP 2007-2013;
- Forecast on the developments in the agricultural and other economic sectors WITH the RDP 2007-2013.

All other parameters have been kept equal. The sectorial forecast was based on OECD and FAO studies, whereas the policy framework was determined by the Health Check Reform of 2009.

Results

The results of the study are depicted in quantitative tables and visualised in coloured maps (NUTS 3) in the annex of the study. A short glimpse on the results of the comparative model simulation reveals:

- Gross Value Added and Employment have increased by a factor 1,4 of the funds invested;
- Regional disparity was reduced;
- Predominantly urban regions have benefited although the funds were mainly invested in rural or integrated regions;
- Beside agriculture, the sectors of construction and engineering as well as service provision to enterprises reap increased value added due to indirect effects;
- The agricultural sector largely benefited; production, employment and farm incomes increased, but gross value added has slightly decreased;
- Pressure on the environment has been alleviated (less nitrogen), and environmentally desirable practices (cultivating marginal land, extensive and organic practices etc.) have been maintained.

The multi-dimensional approach allows for consistent quantitative results which can be qualitatively explained through the hypothesised cross-relationships and results chains.

8.7 The approach consisted in surveying birds' population via ornithologists placed in a network of 168 observation and listening points and 9 linear transects (Veneto, Italy).

Background

The presented case is based on a tailor-made method[185] for assessing the impact on biodiversity of measures 214 (agri-environmental payments) and 216 (non-productive investments) and to calculate the impact indicator no 4 "reversing biodiversity decline", measured by farmland bird species population". The method was implemented in the mid-term evaluation of the RDP of Veneto, Italy. Apart from the evaluator, the team involved ornithologists to conduct the survey of birds' population.

Method

The approach was based on surveying birds' population conducted by ornithologists using the network of 168 observation and listening points and 9 linear transects. In selecting the observation and listening points, the evaluator took the following steps:

1. Analysis of the regional cartography of Veneto[186];
2. Identification of circles of 3,14[187] ha using a GIS software in both AEM treated and non-treated areas (e.g. organic and conventional orchards);
3. Set-up of control groups based on the identified areas;
4. Communication of the geographical position of treated and non-treated areas to ornithologists who could later identify them in the field using GPS or Google Maps.

Observation and listening points were placed both in treated and non-treated areas of at least 3,14 ha.

Ornithologists collected data during windless and clear days, filling a table reporting the watched and heard birds within a range of 100 meters from their observation and listening points. They also gather information on the land use (e.g. orchards, arable lands, etc.).

All in all, 168 birds' observation and listening points were established and located both in areas under intervention and in control areas.

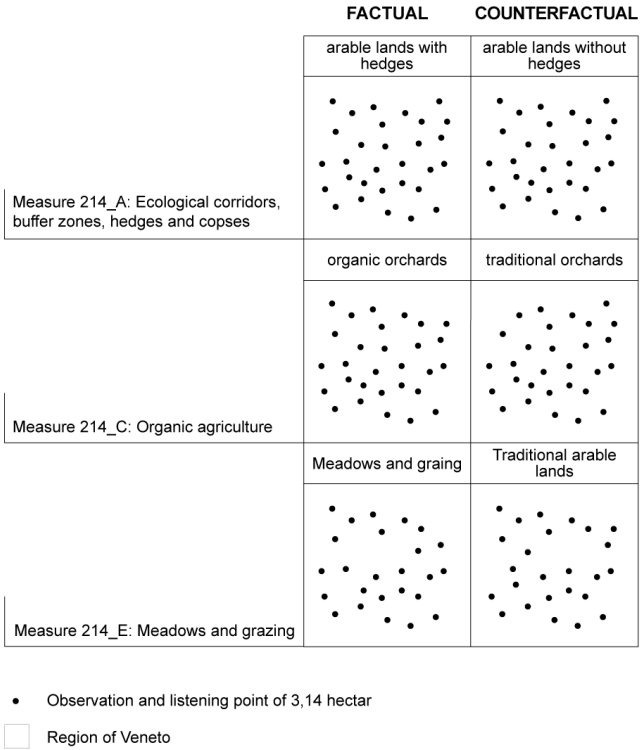
Factual and counterfactual areas were selected among areas relatively close to each other (within 5 Km), at the same altitude and with the same agricultural land use.

The following distinctions were considered when doing the analysis:

- For measure 214, sub-measure A "ecological corridors, buffer zones, hedges and copses", 30 observation and listening points placed in arable land with hedges were paired with 30 observation and listening points placed in arable land without hedges;
- For measure 214, sub-measure C "organic agriculture", 29 observation and listening points placed in organic orchards were paired with 29 observation and listening places placed in conventional orchards;
- For measure 214, sub-measure E "meadows and grazing", 25 observation and listening points placed in meadows and grazing lands were paired with 25 observation and listening points placed in arable land.

[185] Developed by the Italian joint-stock company Agriconsulting
[186] Banca dati dell'uso e copertura del suolo (DUSAF)
[187] The value 3,14 was based on the experience of the MITO2000 project, in which ornithologists suggested to observe bird' s species in a range of 100 meters around of each selected observation point, which represents the area of 3,14

Figure 21 Placement of observation and listening points

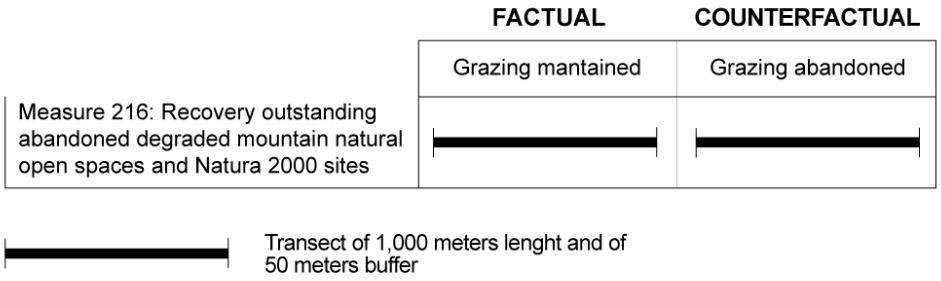


In addition, besides using observation and listening points, also the linear transect method was implemented[188].

The linear transect method was used to assess the impacts of measure 216 (support for non-productive investments) action 6 (recovery of outstanding but abandoned and degraded mountain natural open spaces and Natura 2000 sites) or in other words to estimate the quantity of species located in mountain areas where it was foreseen to abandon grazing for consequent afforestation.

For an observation period of three years (2009, 2010, 2011), at the beginning of the analysis (2009) control groups have been placed in 9 areas where it was foreseen grazing and 9 areas where grazing was foreseen to be abandoned. Linear transects were placed in both areas and ornithologists walked along the transect paths collecting notes on the listened or watched birds within a range of 50 meters along both side of transect. For each species and for both types of area, data was gathered and analysed each 1000 meters of detection.

Figure 22 Linear transect method



Finally, using both data collected from the observation and listening points and the linear transect, two indicators were calculated to assess the impact of the RDP measures on biodiversity.

For a period of three years (2009, 2010 and 2011), for each detecting points and transect, data were gathered and analysed as follow:

[188] Järvinen and Väisänen 1975, Bibby et al. 2000

- The value of 0,5 was given to birds without territorial behaviour[189];
- The value of 1 was given to birds with territorial behaviours (birds with territorial vocalisations, involved in territorial fights or transporting beakfuls, nest material or faecal sacs);
- The value of 0 to migrating species only temporary located in the territory.

Once the survey was conducted, the average of values collected during three years of analysis was calculated, in order to minimise unbalances due to yearly demographic variations of bird species".

Finally, in order to evaluate the impact of the RDP measures on the richness of birds' species population and on the abundance of birds' species of conservation interest, two indicators were calculated.

Richness of species indicator [%]

The first method based on the collected data from listening and observation points and transects, estimate the percentage variation of the richness of birds species between factual and counterfactual areas.

In this case, an increase of the percentage of species richness in the treated areas compared to the non-treated ones indicates a positive effect of the RDP on biodiversity.

In order to obtain the percentage variation of the number of species obtained by the RDP, the following formula was used:

$$\%\, Variations\ of\ richness\ of\ species = \left[\frac{S1a - S2a_{co}}{S3a_{a_{co}}} * 100\right] * pi\ int.a + \left[\frac{S1b - S2b_{co}}{S3b_{b_{co}}} * 100\right] * pi\ int.b + \left[\frac{S1c - S2c_{co}}{S3c_{c_{co}}} * 100\right] * pi\ int.c + \left[\frac{S1d - S2d_{co}}{S3d_{d_{co}}} * 100\right] * pi\ int.d$$

Legend:
- a: Maintenance/construction of meadow;
- b: Maintenance/construction of hedges and groves;
- c: Organic/non organic;
- d: Maintenance of meadow in mountain grazing;
- $S1$: Species richness in the intervention areas;
- $S2$: Species richness in the control areas ($_{co}$);
- $S3$: Overall richness of species between area under intervention and control area (e.g. number of common species of the two areas + number of species present only in the control area);
- co: Counterfactual area;
- $pi\ int.$[190]: is the area under intervention divided by the overall area of the same typology (e.g. meadow) of the region, also taking the altitude into account.

[189]Territorial behaviour, in zoology, is the methods by which an animal, or group of animals, protects its territory from incursions by others of its species. Territorial boundaries may be marked by sounds such as bird song, or scents such as pheromones secreted by the skin glands of many mammals. If such advertisement does not discourage intruders, chases and fighting follow.

[190]The calculation of "pi int.", took also into account the following ratios:
- Area of arable land in a buffer zone of 100 meters around hedges and groves located in a plain divided by the area of arable land in the plain;
- Area of organic orchards in hills and plain divided by the area of orchards in hills and plain;
- Area of maintained meadows in the plain divided by the area of arable land and meadows in the plain;
- Area of maintained meadows and grazing on the mountain divided by the area of grazing and lily of the valley in mountain areas.

Abundance of conservation interest birds species indicator [%]

This second indicator was calculated considering areas which may host a smaller number of species, but have a great ecological value because of nesting sites for species of conservation interest.

Therefore, the following formula was implemented to assess the percentage variation attributable to the RDP on the abundance birds species of conservation interest.

$$\%\, Variations\ of\ abundance\ of\ birds\ species\ of\ conservation\ interest\ =$$
$$= \left[\frac{A1a - A2a_{co}}{A3a_{a_{co}}} * 100\right] * pi\ int.a + \left[\frac{A1b - A2b_{co}}{A3b_{b_{co}}} * 100\right] * pi\ int.b$$
$$+ \left[\frac{A1c - A2c_{co}}{A3b_{c_{co}}} * 100\right] * pi\ int.c + \left[\frac{A1d - A2d_{co}}{A3d_{d_{co}}} * 100\right] * pi\ int.d$$

Legend:
- a: Maintenance/construction of meadow;
- b: Maintenance/construction of hedges and groves;
- c: Organic/non organic;
- d: Lawn maintenance in mountain grazing;
- $A1$: Abundance of individuals in the intervention areas;
- $A2$: Abundance of individuals in the control areas ($_{co}$);
- $A3$: A1 + A2;
- co: Counterfactual area;
- $pi\ int.$: Is calculated as reported for the richness of species indicator.

Results

The outcomes of the study show a significant difference in term of biodiversity between treated and non-treated areas, meaning that the programme positively contributed to preserve or enhance the biodiversity of the region.

The robustness of the analysis is ensured by the time spectrum of the analysis, which is three years.

8.8 Effects of selected Axis 2 measures on carbon sequestration in relation to climate change (Czech Republic)

Background

The study “Effects of selected Axis 2 measures of the 2007-2013 RDP on organic carbon sequestration in relation to climate change was conducted within the frame of ongoing evaluation of RDP Axis 2 measures after several years of their implementation.

The study assesses whether the long-term implementation of above RDP Axis 2 measures (from 1 to 8 years) will mitigate climate change which negatively influences the agriculture production including the development of new types of diseases. The increase of soil and biomass organic carbon supply and concentration, and carbon sequestration are considered the main benefits of the measures implementation in order to mitigate climate change, therefore they can be seen as the programme specific measurement for the impact indicator “combating climate change “.

The method applied in the study is suitable for *ex post* evaluation, when respective measures have been implemented for sufficient length of time to show their effects. Although the study relates to the 2007-2013EU programming period, the applied methodological approach brings important lessons for the evaluation during the programming period of 2014 – 2020, when the Total Soil Organic Carbon (SOC) in arable land expressed in megatons is the unit of measurement for the common impact or context indicator “Soil organic matter in arable land”.

Method

Analyses were conducted on the two sets of parcels – both treated and non-treated in relation to selected RDP Axis 2 measures: (2 AES (M 214) measures : organic farming and conversion of arable land into grass land and measure M 221: afforestation of agriculture land). Altogether 100 parcels of each set were involved in the analysis, 30 parcels for each type of measure and 10 for the assessment

of the waterlogged arable land. Spatial distribution of parcels for practical reasons respected the most important physical-geographical and environmental variables the same way as they exist in the Czech Republic agriculture and has covered all regions of the country.

The databases of the Ministry of Agriculture or publicly accessible databases have been used in the selection of tested parcels:

(1) geographic database LPIS in the Esri shapefile format - „LPIS_to_OEP.shp“ (update of 2010),

(2) database in the Microsoft Access Database format - „Titles AEO2010_export.mdb“ with additional attributes of parcels (update of 2010),

(3) online database LPIS (http://eagri.cz/public/app/lpisext/lpis/verejny/) for waterlogged areas and areas under irrigation/drainage (layers of “Waterlogged land” and “Irrigation/drainage”).

The information about the length of contracts was collected from the paying agency (SZIF) sources provided by the Ministry of Agriculture:

To characterise parcels a set of *ex ante* parameters has been created for the entire Czech territory. In this approach each parcel was marked with a code (category), composed of 3 numbers linked to 3 important characteristics on the parcel environment:

- climatic (obtained from the database BPEJ – Bonita soil-ecological unit) – 1st number in the code;
- chemical composition of soil – content of alkali in the soil substrate – 2nd number in the code;
- relief represented with parcel steepness (obtained from the LPIS database) – 3rd number in the code.

The above approach enabled to generate 120 combinations of characteristics (categories marked with codes). The area of all parcels with the same combination was summed up. Out of this number experts have selected 24 categories with the highest country area coverage, which represent 82% of the total surface of the Czech Republic.

Another *ex ante* parameter was the “length of the contract”. For afforestation and conversion of arable land to grassland the parcels with longest or shortest length of contract have mostly been selected. As for organic farming parcels with longest contracts were only included in the study.

For each parcel the control counterpart was considered as matching if both counterfactual parcels had identical first number of the code (climatic parameter) and differed in:

- -/+1 for other two numbers, or
- -/+1 for the number linked to “steepness” category plus they were located close to each other.

In addition all control parcels had to belong to “arable land” without AES or afforestation contracts.

Each parcel has been divided into three parts structured around the gradient of the steep line of the parcel (low, medium, high part). Samples were taken from the part’ s centroids, which were set up 20 m from the parcel edge. In case of any barriers due to hydrological or morphological anomalies the sample was taken from extra centroid, which was set up together with the previous one.

The following characteristics were noted when collecting samples from parcels:

- Description of the parcel: type and stage of the crop, surface micro-morphology, loosening soil, slope steepness, structure of the forestry vegetation, photo documentation,
- Collection of soil samples and leftovers: for chemical analysis, bulk density and quantification of leftovers.

The following main variables have been applied in the study:

- The length of the application of the measure (afforestation and conversion of the arable land into grassland),
- Type of land management (organic/conventional),
- Type of soil (categories of soil types),
- Hydrology of parcels (waterlogged/non-waterlogged).

In addition the following set of co-variable was tested:

- Elevations above sea level of the parcel centroid (data was obtained via interpolation from the digital model of relief out of contour lines ZABAGED in ArcGIS application).
- Slope steepness of the parcel measured in the field (information was processed from data measured via algorithm used in ArcGIS programme for the calculation of the steepness and exposition - Burrough, McDonell, 1998).
- Slope orientation (information was processed from data measured via algorithm used in ArcGIS programme for the calculation of the steepness and exposition Burrough, McDonell, 1998).
- Climate region (data obtained from database BPEJ[191]).
- Main soil unit (data obtained from database BPEJ).
- Potential content of alkali in soil substrate (data was obtained from the BPEJ code characteristics (Regulation of MoA 327/1998 Sb.)).
- Topographic wetness index – TWI (data obtained from SAGA GIS application, digital model of relief of contour lines ZABAGED, algoritm Quinn et al. 1991, Freeman, 1991 and calculation formula of Kopecký, 2013).
- Average annual rainfall (data obtained from projection of layers of parcel samples on geo-referenced digitalised map of average annual rainfalls in ArcGIS application and deducted with values attributed to the following scale: <450, 450, 500, 550, 600, 650, 700, 800, 1000, 1200 mm).
- Average 6 month summer period rain fall (data obtained from projection of layers of parcel samples on geo-referenced digitalised map of average annual rainfalls in summer 6 months in ArcGIS application and deducted with values attributed to following scale <300, 300, 325, 350, 400, 450, 500, 600, 700 mm).
- Average annual temperature (data obtained from projection of layers of parcel samples on geoferent digitalised map of average annual temperature in ArcGIS application and deducted with 1°C.).
- Average 6 month summer period temperature (data obtained from projection of layers of parcel samples on geoferent digitalised map of average annual temperature in summer 6 months in ArcGIS application and deducted with 1°C.).

[191] BPEJ – Bonita soil-ecological unit

- Lang rain factor (data obtained from projection of layers of parcel samples on geoferent digitalised map of Lang rain factor in ArcGIS application and deducted with values attributed to following scale: < 60, 60, 70, 80, 100.).
- C/N ratio (laboratory tests).
- pH (laboratory test).
- Soil bulk density (laboratory tests).
- Content of skeleton (Data obtained from the BPEJ code characteristics).
- Average sum up of reference evapotranspiration (Data obtained from projection of layers of parcel samples on geoferent digitalised map of average evaporation in summer 6 months in ArcGIS application and deducted with values attributed to the following scale: <450, 450, 500, 550, 600 mm.).

The concentration of carbon in the soil was analysed with the analyser NCS Flash 2000 with auto-sampler (Thermo Scientific, USA). Total an-organic carbon (TIC) was analysed with the analyser CS 500 (ELTRA GmbH, Germany). Further the organic carbon concentration in biomass was analysed using the approach of GHG inventory and the methodology IPCC, which is usually used in the inventory of GHG emissions.

On each of parcels the total organic carbon (TOC) has been measured in the above mentioned 3 soil samples. As for measures: conversion into the grassland and afforestation as well as the concentration of organic carbon in living and death biomass was analysed. Based on these analyses the concentration of the organic carbon per area unit was calculated.

Data obtained from tested parcels and samples as well as from other collected variables have been statistically analysed using the set of descriptive statistical indicators created for individual layers of data (average, median, standard deviation, minimal and maximal values). Mutual co-relations among individual variables were tested with Spearman co-relation coefficient and presented in linear regression. Statistical analysis of obtained data was based on the mixed linear models with mixed effects.

Results

The assessment of effects of selected RDP Axis 2 measures on the organic soil carbon supply enabled to formulate the following conclusions:

- All three measures as well as long-term waterlogged land influence positively carbon sequestration, carbon concentration in the 20 cm layer and supply in relation to reference weight of the soil;
- Afforestation had the highest sequestration potential per time unit (3 Mg C ha-1.year-1) out of the selected measures. Afforestation also causes the continuous long term growth of the carbon supply in the biomass and the share of bounded carbon in total carbon supply per area unit, which can reach the 100% or more of total carbon supply after 8 years of the measure application. In average biomass bounded in the tree is 14 Mg C per ha after 8 years of measure application;
- In case of conversion of the arable land into the grassland the highest increase of carbon supply was observed during the first years after the measure application, however it was stabilised in later stages;
- For both measures (afforestation and conversion into grassland) the length of implementation was an important factor;

- The measure organic farming had shown the lowest carbon sequestration potential – in average 0,3 Mg C ha-1.rok-1, however the time co-relation of carbon sequestration was not assessed for this measure. On the other hand the relation between carbon sequestration and soil type was statistically significant. Floodplain and gleyic soil reached the double supply of organic carbon compare to cambisoils and brawn soils.
- Waterlogged arable land had shown a high carbon sequestration potential - 16,2 Mg C ha-1 compared to control plots.
- Carbon concentration and supply is in negative co-relation with the soil bulk density;
- The estimation of sequestrated carbon in 2012 for conversion of arable land into grassland was 742 043 Mg, in case of afforestation - 146 817Mg, in case of organic farming - 51 689 Mg, and altogether 940 549 Mg of carbon, which represent 3 448 679 Mg CO2.

8.9 Propensity score matching

Method

The main steps involved in carrying out a binary propensity score matching (PSM) are as follows:

(a) Obtain a dataset including information on basic characteristics and performance of programme beneficiaries and non-beneficiaries in two time periods, i.e. prior to implementation of a given RDP and after. One possibility to pre-select individual units of programme non-beneficiaries to a dataset is to apply programme or measure eligibility criteria.

(b) Compute differences in all basic characteristics and performance of programme beneficiaries and non-beneficiaries prior to applying matching.

(c) Run a participation model (probit/logit regression). Generally, covariates entering the probit/logit function are expected to determine both programme participation and outcomes (the latter are typically measured in terms of relevant result indicators at micro level).

(d) Calculate participation probabilities for each individual unit (programme beneficiaries and non-beneficiaries) included in the dataset.

(e) Drop observations outside the region of common support (i.e. individual observations in the group of programme beneficiaries whose probability of receiving support exceeds that of any from the potential comparison group, or those from the control group with probabilities of receiving programme support below those of any members of the group of programme beneficiaries).

(f) Match observations based on participation probabilities (here various matching algorithms can be applied. Selection of appropriate matching algorithm should be a subject to statistical analysis, e.g. by i) applying % of the standardised bias reduction – after matching -; or ii) applying pseudo R^2 test after matching - as a selection criterion).

(g) Calculate programme results for each pair or set of matched observations.

(h) Calculate the average of these differences for a period prior to and after programme implementation using ATT combined with DID.

(i) Perform sensitivity analysis.

Table 22 Comparison of a binary PSM with a generalised PSM approach

Binary Propensity Score Matching			
Main hypotheses	**Data required**	**Pros**	**Cons**
Programme selection is determined only by observable characteristics not influenced by programme support (yet, proxy variables can be included as controls for unobservables).	The dataset contains all the variables describing major characteristics and performance of programme beneficiaries and non-beneficiaries prior and after implementation of the programme.	• Very effective tool applied in impact evaluation of various programmes. • Very effective tool for finding counterfactuals. • It controls a set of covariates that simultaneously determine the decision to participate in a given programme. • Is a non-parametric approach, therefore very flexible (i.e. matching does not require any functional form assumptions for relationship linking the outcome variable with the covariates).	• It can only be performed on observed characteristics. • It will not generate reasonable results if other important observable characteristics which explain differences between programme beneficiaries and control group were not included in the model. • It is a data-hungry procedure. • Its conclusions hold only on the subset of matched units. • The external validity of its results decreases when the share of unmatched units increases.
Generalized Propensity Score Matching			
Main hypotheses	**Data required**	**Pros**	**Cons**
Programme selection is determined only by observable characteristics not influenced by programme support (yet, proxy variables can be included as controls for unobservables).	The dataset contains all the variables describing major characteristics and performance of programme beneficiaries prior and after implementation of the programme, and intensity of programme support obtained during implementation of RDP	• It does not require existing units that did not receive programme support; • Very effective tool applied in impact evaluation of various programmes if almost all units receive a programme support (at various intensity levels). • Very effective tool for finding counterfactuals. • It controls a set of covariates that simultaneously determine the obtained intensity level of programme support. • It allows to assess the result/impact of a given programme for each relative level (between 0% to 100%) of obtained support	• It will not generate reasonable results if other important observable characteristics which explain differences between various programme beneficiaries were not explicitly included in the model. • It requires high analytical skills.

9 TOOL TO IDENTIFY DATA SOURCES FOR EVALUATION

The table below intends to support MAs and evaluators in identifying data sources for RDP evaluation, in particular for evaluation methods applying counterfactual. Once the table is filled with all known data sources, it shows where data gaps occur and which need to be bridged, either by the MA or the evaluator.

The table is subdivided in five categories according to the potential RDP beneficiaries: 1) agricultural holdings, 2) forestry holdings, 3) food holdings, 4) villages and, 5) micro-enterprises, NGOs. The required data sets should always cover the total population; therefore include RDP beneficiaries and non-beneficiaries.

The data listed in the tables is not exhaustive and may not cover all areas needed for specific RDP evaluation. Therefore MAs are encouraged to add items or change the list according to their specific evaluation needs or methods employed.

For all the data sources, it is advisable to collect the following information in order to have complete and quality full data sets:

- Description of data
- Population size (number of units for which data are collected)
- Owner
- Web link (URL)
- Measurement unit of data (e.g. EUR, tons, kg/ha)
- Availability of database(e.g. on the web for free; on the web for purchase; available on request (free); available on request (purchase); not available)
- Name of database
- Remarks

In the tables below the data source table is filled out for the Slovak Republic, the first four columns are presented.

Agriculture

Agricultural Farms (total population – beneficiaries and non-beneficiaries)	**Population size**	**Owner**	**Weblink (URL)**
Data on obtained level of support from each RDP measure (by measure and total)	national, NUTS III, IV	Agricultural Paying Agency (APA) *(remark: available partially (investment measures on national level or NUTS III))*	http://www.apa.sk/index.php?navID=353
Data on obtained level of support from other EU and national programmes (total)	national	CCA - coordination of programming on national level *(remark: each support programme has its own web site and own monitoring/ reports)*	http://www.nsrr.sk/operacne-programy/
Main enterprise/holding structural characteristics, e.g. including:			
Production (value and structure, incl. crops(*), livestock(*); marketable and non-marketable; high quality value, etc.); (*)Agricultural Farms only	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Turnover/Revenues (level and structure)	national, NUTS III	National Statistical Office	www.statistics.sk
GVA (see: http://enrd.ec.europa.eu/app_templates/filedownload.cfm?id=84053593-C697-FF89-ED5C-51797D9754FD; and http://enrd.ec.europa.eu/app_templates/filedownload.cfm?id=8D7932B6-D3A1-95CC-8A26-14ABFAB6D327)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
The level and (%) structure of non-agricultural GVA (e.g. tourist vs. non-tourist origin)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Profits and Gross Margins	national, NUTS III	National Statistical Office	www.statistics.sk
Household income (agricultural and non-agricultural)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Area (structure of land ownership; and land use structure; of which NATURA2000 sites and river basin areas affected by the WFD; first afforestation, etc.)		National Statistical Office, CD Ministry of Agriculture and Regional Development (MADR) of the SR, National Agriculture and Food Center (NAFC) -Research institute for agriculture and food economics (RIAFE), Water Research Institute (WRI) , National Forest Center (NFC) SR, NAFC - Soil Science and Conservation Research Institute, Eurostat	

Agricultural Farms (total population – beneficiaries and non-beneficiaries)	Population size	Owner	Weblink (URL)
Employment (i.e. agricultural, non-agricultural; family, non-family; educational level; age structure; etc.)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Labour costs (hired labour)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Value and structure of capital (buildings, machinery, livestock)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Value of investments	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Milk quotas, etc.		Agricultural Paying Agency	
Yields (for Agr.Farms by main crops and livestock)	national, NUTS III, IV	National Statistical Office, MADR	www.statistics.sk
Use of fertilizers		The Central and Testing Institute in agriculture in Bratislava, Slovakia (ÚKSÚP)	
Purchased feed		National Statistical Office, MADR	www.statistics.sk
Total costs		National Statistical Office	www.statistics.sk
Data on area under successful land management contributing to:			
biodiversity and High Nature Value farming/forestry		NAFC - Soil Science and Conservation Research Institute + National Forest Centre (NFC) Ministry of environment/State nature protection of SR MADR SR	
protection of wildlife species or groups of species		National Forest Centre, Ministry of Environment/State nature protection of SR	
safeguarding endangered animal breeds and plant varieties		NAFC-Plant Production Research Institute; NAFC-Animal Production Research Centre, Slovak University of Agriculture	
land cover types corresponding to HNV farmland (i.e. semi-natural pastures and meadows; traditional orchards; mosaics of low-intensity crop types; fallow land in low intensity farming systems; natural and semi-natural forests);		NAFC - Soil Science and Conservation Research Institute, NAFC - Plant Production Research Institute, National Forestry Center MADR SR	

Agricultural Farms (total population – beneficiaries and non-beneficiaries)	Population size	Owner	Weblink (URL)
farming practices supporting HNV land (low livestock densities, low fertiliser and pesticide input, lower yields compared to regional averages, presence of understorey in permanent crops);		NAFC-Plant Production Research Institute; NAFC-Animal Production Research Centre, MADR	
water quality			
decrease in concentration of nutrients, phosphorous and/or pesticides		NAFC - Soil Science and Conservation Research Institute; The Central and Testing Institute in agriculture in Bratislava, Slovakia (ÚKSÚP); Water Research Institute (WRI) Slovak hydrological institute (SHI)	
reduced use of chemical fertilizers		NAFC - Soil Science and Conservation Research Institute; The Central and Testing Institute in agriculture in Bratislava, Slovakia (ÚKSÚP); Water Research Institute (WRI)	
reduced life stock density		NAFC - Soil Science and Conservation Research Institute, The Central and Testing Institute in agriculture in Bratislava, Slovakia (ÚKSÚP)	
improved nitrogen balance		NAFC - Soil Science and Conservation Research Institute; The Central and Testing Institute in agriculture in Bratislava, Slovakia (ÚKSÚP)	
reducing the transport of pollutants to aquifers		NAFC - Soil Science and Conservation Research Institute	
mitigating climate change		Slovak hydrometeorological institute (SHMÚ)	http://www.shmu.sk/sk/?page=1817
soil quality		NAFC - Soil Science and Conservation Research Institute	
reduction of erosion (water/wind/tillage)		NAFC - Soil Science and Conservation Research Institute	
less water logging		Water Research Institute (WRI); NAFC - Soil Science and Conservation Research Institute	
reduction or prevention of chemical contamination (less use of plantnutrient/manure, plant protection substances, …)		NAFC - Soil Science and Conservation Research Institute; The Central and Testing Institute in agriculture in Bratislava, Slovakia (ÚKSÚP)	

Agricultural Farms (total population – beneficiaries and non-beneficiaries)	Population size	Owner	Weblink (URL)
stabilising and enhancing the level of soil organic matter through the use of appropriate sources of stable organic matter and, where appropriate, through reduced tillage		NAFC - Soil Science and Conservation Research Institute	
avoidance of marginalization and land abandonment		NAFC - Soil Science and Conservation Research Institute	
Data on price received for selected high quality outputs		National Statistical Office, Eurostat (EU level)	www.statistics.sk, http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes
Data on value of agricultural/forestry/food production under recognized quality standards (EU and national)		National Statistical Office, Eurostat (EU level)	www.statistics.sk, http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes
Data on environmental conditions at a enterprise/holding level:		*(remark: here we are looking for particular surveys carried out for collecting environmental information linked to the enterprise level (so not the area linked information))*	
Soil erosion	national	NAFC - Soil Science and Conservation Research Institute/	
Water quality	national	primary EU source: DG Agri, European Union NAFC - Soil Science and Conservation Research Institute, Water Research Institute (WRI)	http://ec.europa.eu/agriculture/statistics/rural-development/2013/full-text_en.pdf
Nutrient and pest management	national	The Central and Testing Institute in agriculture in Bratislava, Slovakia (ÚKSÚP); or National Statistical Office	
Air pollution - Greenhouse Gas emissions by sector, etc.	national	Eurostat	http://epp.eurostat.ec.europa.eu/portal/page/portal/environment/data/main_tables
Data on public health conditions at an enterprise/holding level	national	Public Health Authority of the SR); Ministry of Health of the SR	http://www.uvzsr.sk/; http://www.health.gov.sk/Titulka

Agricultural Farms (total population – beneficiaries and non-beneficiaries)	Population size	Owner	Weblink (URL)
Data on animal and plant health conditions at a enterprise/holding level	national	State Veterinary and Food Administration of the Slovak Republic, The Central and Testing Institute in agriculture in Bratislava, Slovakia (ÚKSÚP), MARD of the SR, Slovak Agricultural and Food Chamber	http://www.svssr.sk/, www.uksup.sk, www.sppk.sk,
Data on occupational safety conditions at an enterprise/holding level	national	Ministry of Labour, Social Affairs and Family of the SR; - Central Office of Labour, Social Affairs and Family	http://www.upsvar.sk/
number of occupational injuries and diseases	national	Ministry of Labour, Social Affairs and Family of the SR; - Central Office of Labour, Social Affairs and Family	http://www.upsvar.sk/
Data on hygiene and animal welfare conditions at an enterprise/holding level (if applicable	national	State Veterinary and Food Administration of the Slovak Republic, MARD of the SR, Slovak Agricultural and Food Chamber, unions/associations	http://www.svssr.sk/, www.mpsr.sk, www.sppk.sk,
Conclusions from animal welfare control reports by inspection service	national	State Veterinary and Food Administration of the Slovak Republic, MARD of the SR, Slovak Agricultural and Food Chamber, unions/associations	http://www.svssr.sk/, www.mpsr.sk, www.sppk.sk,
Data on animal welfare conditions which go beyond the relevant mandatory standards	national	State Veterinary and Food Administration of the Slovak Republic, MARD of the SR, Slovak Agricultural and Food Chamber, unions/associations	http://www.svssr.sk/, www.mpsr.sk, www.sppk.sk,

Forestry

Forestry holdings (total population – beneficiaries and non-beneficiaries)	Population size	Owner	Weblink (URL)
Data on obtained level of support from each RDP measure (by measure and total)	national, NUTS III, IV	Agricultural Paying Agency	
Data on obtained level of support from other EU and national programmes (total)	national	CKO (Central Coordination Authority) - coordination of programming on national level	
Main enterprise/holding structural characteristics, e.g. including.:			
Ownership or use of forests		National Forest Centre	www.forestportal.sk
Forest area for timber production		National Forest Centre	www.forestportal.sk
Timber feeling (structure):		National Forest Centre	
* total		National Forest Centre	
* incidental		National Forest Centre	
Forest regeneration (total):		National Forest Centre	
* artificial regeneration		National Forest Centre	
* natural regeneration		National Forest Centre	
Cleaning		National Forest Centre	
Thinning		National Forest Centre	
Forest road network		National Forest Centre	
Forest certification		Slovak Forest Certification Association	
Earnings and revenue		National Forest Centre	
Timber sale revenue		National Forest Centre	
Data on area under successful land management contributing to:			
biodiversity and High Nature Value farming/forestry		NAFC - Soil Science and Conservation Research Institute + National Forest Centre	
protection of wildlife species or groups of species		National Forest Centre	
safeguarding endangered animal breeds and plant varieties			
land cover types corresponding to HNV farmland (i.e. semi-natural pastures and meadows; traditional orchards; mosaics of low-intensity crop types; fallow land in low intensity farming systems; natural and semi-natural forests);			
farming practices supporting HNV land (low livestock densities, low fertiliser and pesticide input, lower yields compared to regional averages, presence of understorey in permanent crops);			

Forestry holdings (total population – beneficiaries and non-beneficiaries)	Population size	Owner	Weblink (URL)
maintenance of High Value forest eco-system		National Forest Centre	
% damaged forest land in total forest land area		National Forest Centre	
water quality			
mitigating climate change			
soil quality			
less water logging			
reduction or prevention of chemical contamination (less use of plant nutrient/manure, plant protection substances, …)			
Data on quality standards/norms at a enterprise/holding level			
Data reflecting applied production techniques (% of environmentally friendly techniques) at an enterprise level			
Data on new technology and innovation in processing of products			
Data on price received for selected high quality outputs		SO of the SR, Eurostat (EU level)	www.statistics.sk, http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes
Data on value of agricultural/forestry/food production under recognized quality standards (EU and national)		SO of the SR, Eurostat (EU level)	www.statistics.sk, http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes
Data on occupational safety conditions and standards (e.g. reports of inspection services) at an enterprise/holding level			
Data on environmental conditions at a enterprise/holding level:		*(remark: here we are looking for particular surveys carried out for collecting environmental information linked to the enterprise level (so not the area linked information)*	
Soil erosion			
Water quality			
Nutrient and pest management			
Air pollution			
other…			
Data on public health conditions at an enterprise/holding level			
Data on occupational safety conditions at an enterprise/holding level			
number of occupational injuries and diseases			

Food

Food processors (total population – beneficiaries and non-beneficiaries)	**Population size**	**Owner**	**Weblink (URL)**
Data on obtained level of support from each RDP measure (by measure and total)	national, NUTS III, IV	Agricultural Paying Agency	http://www.apa.sk/index.php?navID=353
Data on obtained level of support from other EU and national programmes (total)	national	CCA - coordination of programming on national level	http://www.nsrr.sk/operacne-programy/
Main enterprise/holding structural characteristics, e.g. including.:			
Production (value and structure, incl. crops(*), livestock(*); marketable and non-marketable; high quality value, etc.); (*)Agricultural Farms only	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Turnover/Revenues (level and structure)	national, NUTS III	National Statistical Office	www.statistics.sk
GVA (see: http://enrd.ec.europa.eu/app_templates/filedownload.cfm?id=84053593-C697-FF89-ED5C-51797D9754FD; and http://enrd.ec.europa.eu/app_templates/filedownload.cfm?id=8D7932B6-D3A1-95CC-8A26-14ABFAB6D327)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
The level and (%) structure of non-agricultural GVA (e.g. tourist vs. non-tourist origin)	national, NUTS III	National Statistical Office	www.statistics.sk
Profits and Gross Margins		National Statistical Office	www.statistics.sk
Household income (agricultural and non-agricultural)	national, NUTS III, IV		www.statistics.sk
Employment (i.e. agricultural, non-agricultural; family, non-family; educational level; age structure; etc.)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Labour costs (hired labour)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Value and structure of capital (buildings, machinery, livestock)	national, NUTS III, IV	National Statistical Office	www.statistics.sk
Value of investments	national, NUTS III	National Statistical Office	www.statistics.sk
Total costs	national, NUTS III	National Statistical Office	www.statistics.sk
Technical efficiency coefficients			

Food processors (total population – beneficiaries and non-beneficiaries)	**Population size**	**Owner**	**Weblink (URL)**
Data on price received for selected high quality outputs		National Statistical Office, Eurostat (EU level)	www.statistics.sk, http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes
Data on value of agricultural/forestry/food production under recognized quality standards (EU and national)		National Statistical Office, Eurostat (EU level)	www.statistics.sk, http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes
Data on environmental conditions at a enterprise/holding level:		*(remark: here we are looking for particular surveys carried out for collecting environmental information linked to the enterprise level (so not the area linked information))*	
Air pollution			

Villages

Local communities/ vilages, etc. (total population – beneficiaries and non-beneficiaries)	Population size	Owner	Weblink (URL)
Data on obtained level of support from each RDP measure (by measure and total)	national, NUTS III, IV	Agriculture paying agency	
Data on obtained level of support from other EU and national programmes (total)	national	CCA (Central Coordination Authority) - coordination of programming on national level	http://www.nsrr.sk/operacne-programy/
Main structural characteristics of community/village			
Area	NUTS IV	National Statistical Office	http://www.statistics.sk
Natural resources			
Level of urbanisation	NUTS IV	National Statistical Office	http://www.statistics.sk
Population (level and structure)	national, NUTS III, IV, V	National Statistical Office	http://www.statistics.sk
Main economic branches (structure, incl. tourism and agro-tourism)	national, NUTS III, IV	National Statistical Office	http://www.statistics.sk
Personal incomes (total and structure)	national, NUTS III, IV	National Statistical Office	http://www.statistics.sk
Employment (level and structure by branches) and unemployment (level and structure)	national, NUTS III	National Statistical Office	http://www.statistics.sk
Migrations (in- and out-migrations)	national, NUTS III, IV	National Statistical Office	http://www.statistics.sk
Social capital (incl. cultural infrastructure)	national, NUTS III, IV	National Statistical Office	http://www.statistics.sk
Traffic infrastructure	national, NUTS III, IV, V	National Statistical Office	http://www.statistics.sk
Housing	national, NUTS III, IV, V	National Statistical Office	http://www.statistics.sk
Markets infrastructure	national, NUTS III	National Statistical Office	http://www.statistics.sk
Health conditions	national, NUTS III, IV, V	National Statistical Office	http://www.statistics.sk
Educational infrastructure	national, NUTS III	National Statistical Office	http://www.statistics.sk
Environmental indicators	national, NUTS IV	National Statistical Office	http://www.statistics.sk
Other specific quality of life indicators	national, NUTS III, IV, V	National Statistical Office	http://www.statistics.sk
other…			

Micro-enterprises, NGOs

Micro-enterprises, NGOs, etc. (total population – beneficiaries and non-beneficiaries)	Population size	Owner	Weblink (URL)
Data on obtained level of support from each RDP measure (by measure and total)	national, NUTS III, IV	Agriculture paying agency	
Data on obtained level of support from other EU and national programmes (total)	national level - SR	CCA (Central Coordination Authority) - coordination of programming on national level	
Main structural characteristics of community/village			
Turnover/Revenues (level and structure, incl. agriculture and non-agriculture)	national level, NUTS III, IV	National Statistical Office	http://www.statistics.sk
Profits			
Source of income (agricultural and non-agricultural)	2.600 farms	MADR SR (remark: for farmers)	
Employment (i.e. agricultural, non-agricultural; family, non-family; educational level; age structure; etc.)	national, NUTS III	National Statistical Office	http://www.statistics.sk
Labour costs (hired labour)	national, NUTS III and IV	National Statistical Office	http://www.statistics.sk
Value and structure of capital (buildings, machinery, etc.)	national, NUTS III, IV and categories of economic activities	National Statistical Office	http://www.statistics.sk
Total costs	national, NUTS III, IV and categories of economic activities	National Statistical Office	http://www.statistics.sk
Participation in networks			

European Evaluation Network
for Rural Development

Evaluation Helpdesk
260 Chaussée St Pierre – B-1040 Brussels
Tel: +32 (0)2 736 18 90
E-mail: info@ruralevaluation.eu
http://enrd.ec.europa.eu/evaluation