GUIDELINES

ASSESSING RDP ACHIEVEMENTS AND IMPACTS IN 2019

PART IV - TECHNICAL ANNEX

AUGUST 2018

The Evaluation Helpdesk is responsible for the evaluation function within the European Network for Rural Development (ENRD) by providing guidance on the evaluation of RDPs and policies falling under the remit and guidance of DG AGRI's Unit C.4 'Monitoring and evaluation' of the European Commission (EC). In order to improve the evaluation of EU rural development policy the Evaluation Helpdesk supports all evaluation stakeholders, in particular DG AGRI, national authorities, RDP managing authorities and evaluators, through the development and dissemination of appropriate methodologies and tools; the collection and exchange of good practices; capacity building, and communicating with network members on evaluation related topics.

Additional information about the activities of European Evaluation Helpdesk for Rural Development is available on the Internet through the Europa server (http://enrd.ec.europa.eu).

GUIDELINES

ASSESSING RDP ACHIEVEMENTS AND IMPACTS IN 2019

PART IV - TECHNICAL ANNEX

AUGUST 2018

CONTENT

TABLES AND FIGURES

4   TECHNICAL ANNEX

PART IV (Technical Annex) of the non-binding Guidelines 'Assessing RDP achievements and impacts in 2019' contains complementary information for the evaluation approaches proposed in Part II, examples of additional indicators, detailed descriptions of working steps, adequateness of suggested evaluation approaches, dos and don'ts and a glossary. It is to be read in context with **PART I** (informing Managing Authorities about the legal requirements of the AIR submitted in 2019), with **PART II** (methodological support to evaluators for assessing the common impact indicators of Pilar II) and **PART III** (Fiches for answering the CEQs 22-30).

4.1   CAP common impact indicators related to the agricultural sector: I.01, I.02, I.03

The following sections of the Technical Annex are related to Chapter 2.2 in PART II of the Guidelines.

4.1.1   Additional indicators (examples)

Additional indicators might be used to assess the indicators-constituents of farm income and productivity. They might include GVA-, agricultural production per labour unit (AWU) or additional partial productivities (in respect to land and capital). Furthermore, it is useful to include indicators of profitability, efficiency and financial stability like cost/revenue ratios, technical and allocative efficiency indicators, return to capital ratios or indebtedness ratios.[1] Among other indicators that might be applied are: comparative advantage indicators, indices of commercialisation level, etc. (see: OECD, 2010). In order to establish a robust causal relationship between the RDP and policy induced changes in agricultural competitiveness the evaluators may also use the above indicators calculated at the commodity levels[2] and perform at this level separate sub-evaluations. Additional competitiveness aspects, such as product differentiation, product and service quality and variety, design, novelty, reputation and reliability may also need to be taken into account. Furthermore, the application of additional indicators should lead to a more complete- or less biased picture of the impact of RDP on agricultural competitiveness, e.g. via assigning particular importance to the issue of unpaid inputs and subsidies received.[3]

While common impact indicators (I.01, I.02) concern only income from agriculture activities, it is obvious that the competitiveness of many agriculture holdings and consequently the agriculture sector depends also on non-agricultural income, which is generated on the farm such as renting of farm equipment, rural and agritourism activities etc. FADN variable: SE256 (other output) includes certain types of non-agriculture activities, such as renting equipment or receipts from tourism, which contribute to a variable 'family farm income' SE430, and 'farm net value added' SE425. Should additional farm income data, e.g. from total non-agriculture activities (off-farm income) be available (e.g. from farm bookkeeping records) such data can be used in order to compute the total farm income (from agricultural and non-

---

[1] Particularly, bank credit indebtedness might be useful to see if the intervention stimulates more external capital but keeping in mind that excessive indebtedness might threaten the financial stability of agricultural business.

[2] In the view of some authors, e.g. Capalbo et al. (1990), competitiveness should be measured at the commodity level rather than at the sector level. Finally, some authors claim that measuring a nation's or a sector's competitiveness is meaningless and what matters is individual (firms or farms) competitiveness (e.g. Brinkman, 1987; Krugman, 1994; Harrison and Kennedy, 1997) see: OECD, 2010.

[3] According to Brinkman (1987), government intervention may superficially change competitiveness without increasing real competitiveness. The author explains that in cases where competitiveness is 'bought' by public subsidies, it may be a false competitiveness; see: OECD, 2010.

agricultural activities) which then can be used as an additional indicator of farm competitiveness. Some selection of additional indicators, extending the magnitude of income (and therefore the measure of competitiveness) by including e.g. SE256 'other output', are displayed in Table 1.

The evaluator decides if to include or not additional indicators, particularly those also taking into account on-farm non-agricultural income, depending on the availability of the underlying data, cost of their collection and time.

Table 1. Examples of additional indicators for the measurement of farm competitiveness (based on FADN)

Indicator	Measurement Unit	Use of additional indicator(s)	Data sources and frequency of collection
Family farm income per family work unit = Family Farm Income/FWU	EUR/person	The indicator shows the level of family farm income per family unit. It takes into account differences in the family labour in terms of remuneration	FADN: variable: SE430 = SE420/SE015 Where: SE430 = family farm income per family work unit SE420 = Family farm income SE015 = unpaid labour input (generally equivalent of family work unit = FWU) Collected regularly once per year
Farm net value added per Annual Work Unit = Farm net value added/AWU	EUR/person	The indicator shows the level of remuneration of the fixed factors of production per labour force employed in agricultural activities	FADN: variable: SE425 Where: SE425 = farm net value added per annual work unit Collected regularly once per year
Total output per work unit Total Output/AWU	EUR/person	Shows labour agricultural productivity (=partial factor productivity)	FADN: calculation: SE131/SE010 Where: SE131 = Total Output SE010 = Total labour input Collected regularly once per year
Total output per unit of land Total Output/land area	EUR/ha	Shows land agricultural productivity (=partial factor productivity)	FADN: calculation: SE131/SE025 Where: SE131 = total output SE025 = total utilized agricultural area Collected regularly once per year
Costs as % of output	%	Shows farms competitive position without subsidies (and taxes)	FADN: calculation: SE270/SE131 in % Where: SE270 = total inputs (costs) SE131 = Total output Costs = specific costs + overhead costs + factor (land, labour, capital) opportunity costs (e.g. depreciation + external factors) Collected regularly once per year
Subsidies as % of farm net income	%	Shows importance of subsidies in farm net income	FADN: calculation:

Indicator	Measurement Unit	Use of additional indicator(s)	Data sources and frequency of collection
			(SE605 + SE406 + SE407)/SE420 Where: SE605 = total subsidies excl. on investment (incl. total support for RD) SE406 = subsidies on investment SE407 = payment to dairy outgoes SE420 = farm net income Collected regularly once per year
Yields of major agric. products and various productivities	T/ha	Measures various partial productivities	FADN: calculation: e.g. SE110 = yield of wheat SE115 = yield of maize SE125 = milk yield SE132 = total output / total input (includes price effects) Collected regularly once per year

4.1.2 Approach A

Micro-level assessment

Access to data and its quality & creation of consistent database and data infrastructure

The use of existing data is highly recommended for the evaluation exercise in 2019. It is especially proposed to help build the micro level analysis on data from FADN, enriched by the information on participation in RDP measures from the information system of the managing and paying authorities (data on the implementation of measures). The advantage of FADN rests in the consistency of collection and processing of data over years. If a similar farm data survey exists, the evaluator may consider it.

The approach A (PSM-DiD) is dependent on the availability of panel data for the investigated period. Panel data requires replication of the same units over time: ideally prior to and after the implementation of the given measures of a RDP. For the FADN data it might be a challenge if a long period is to be analysed, since farms in the survey may be regularly replaced by new respondents (within a range of up to 15% per year). However, for 2013-2017 (most recent data will not be available in the FADN) it will constitute only a marginal problem.

There should be no problem with the quality of data in FADN in terms of completeness and time consistency, since a sophisticated quality check is done regularly. With alternative databases this might be a challenge and the evaluator has to be ready to deal with it. If the evaluator also needs to take into consideration the effects on the smallest farms (i.e. under the threshold levels applicable to FADN farms, in case those small units have been included in RDP eligibility criteria) some additional surveys for this group would be necessary.

Collected data should adequately describe the structure and performance of analysed units (farms). The biggest part of per unit information (approx. 80%) relates usually to the data block 'structure' which is used to construct meaningful control groups (e.g. via the application of matching techniques, etc.). The second largest part concerns result and impact indicators: the data block 'performance'. The micro-level data may consist of bookkeeping data or survey data (or both). Collected data should concern

programme beneficiaries and non-beneficiaries; they should clearly identify programme beneficiaries and the level of support they received from individual RD measures. Ideally, the micro-level data panel should comprise not less than 150 beneficiaries (farms) and 2-3 times more non-beneficiaries.[4]

It is sometimes reasonable to work with sub-samples of more homogenous farms e.g. dairy farmers or pastoral farming. It is particularly relevant if such groups get differentiated (preferential) treatment within a measure.'

Apart from quantitative data, also qualitative information shall be collected by the evaluator to triangulate quantitative findings or to fill the data gaps. This can be done via using interviews, focus groups, case studies, etc. As in the case of quantitative data collection, it is important to:

- ensure the utility/relevance of the collected qualitative information to answer the EQs (check judgement criteria); and
- ensure that qualitative information collected from beneficiaries can be paired with the information collected from non-beneficiaries to ensure the counterfactual and netting out of the programme effects.

Selection of counterfactual option and micro-level method

In the case of sector indicators, the construction of an appropriate control group requires performing the **six steps** as described in Chapter 2.1 of PART II of the Guidelines (see box Quick Guide #2).

The proposed approach (PSM with DiD) assumes that the evaluator has enough observations on non-supported farms from which a control group can be constructed.[5] If this is not the case, and if the majority of farms included in the FADN (or farm bookkeeping) database obtained support from the RDP (2014-2020), it may happen that due to insufficient number of observations on non-supported farms a binary Propensity Score Matching analysis cannot be applied. In such a situation the direct and indirect effects of the RDP measures on the three impact indicators (I.01, I.02 and I.03) can be alternatively analysed by means of a dose-response function and a derivative dose-response function, see: Generalised Propensity Score Matching (GPSM).[6] Generally speaking, the GPSM method which is also based on a concept of counterfactuals not only allows to estimate the average effect of support from the RDP on selected result/impact indicator (e.g. AFIM), but also to assess the *marginal* effects of the programme 'on all three sectoral impact indicators' depending *on the support intensity level* obtained by each farm from RDP in the current programming period, e.g. in years 2014-2018.[7]

Net impact assessment at micro level

Below we present an approach which can be applied for the calculation of programme net effects on the impact indicators (I.01, I.02 and I.03). It consists of five principal steps, which are generally described in PART II, Chapter 2.1, box Quick Guide #4 'How to assess RDP net effects?'. In addition,

---

[4] While this relation should be treated as optimal, in practice, ratios of beneficiaries to non-beneficiaries of 1:1 or smaller may work quite well (depending on the matching algorithm applied).

[5] In the smallest regions (except Martinique, Guadeloupe, etc.) the coverage of FADN data amounts to approximately 200 farms. This number is usually sufficient to run PSM (for a base year) in order to find out appropriate controls. Should a yearly replacement of FADN farms in this sample lead to a drastic shrink of number of farms remaining on the panel, comparisons of results between periods (DiD) would have to be applied only for farms remaining on the panel. Further information about coverage of FADN can be obtained from Annex II (Article 2) of Commission Implementing Regulation (EU) No 2015/220 of 3 February 2015.

[6] Examples for the assessment of programme net-effects in situation when support is provided to almost all farms are: 'Capitalisation of CAP Single Payment Scheme into Land Value: Generalized Propensity Score Evidence from the EU', Land Economics, University of Wisconsin Press vol. 90, NNr 2, pp: 260-289 http://le.uwpress.org/content/90/2/260.full.pdf+htm, .Michalek J., P. Ciaian and d'A. Kancs, (May 2014).

[7] The detailed steps, i.e. application of a binary PSM-DiD method, are described in the Guidelines for the ex-post evaluation of 2007-2013 RDP; Chapter 4.3.3.2, and for application of GPSM in Hirano and Imbens, 2004; Michalek 2012; Kluve et.al. 2009.

the specific issues as linked to the CAP impact indicators for agriculture sector are highlighted here for Steps 3 - 5:

Step 1: *Estimation of RDP direct effects on supported units at a micro-level* (described in Chapter 2.1).

Step 2: *Estimation of RDP indirect effects on supported units at a micro-level* (described in Chapter 2.1).

Step 3: *Calculation of indirect effects on non-supported units at a micro-level. Calculation of indirect effects* of RDP support on non-supported farms.

For a preliminary analysis of the potential effects, the evaluators may also use the qualitative assessment tool developed in previous guidelines.[8]

Step 4: *Aggregation of results and calculation of RDP effects at a macro-, i.e. programming area level*

This step should not be applied to indicators computed per AWU (both in supported- and non-supported farms). Since the calculated net direct effect of the RDP on the indicator per AWU at a farm level is already expressed as a ratio, the value of the net indicator showing corresponding results per AWU working unit at a programme area level would need to be computed as a weighted average of indicator/AWU (% supported in total) and indicator/AWU (% non-supported in total).

Step 5: *Application of qualitative methods involving triangulation for checking and verifying obtained results*

Given the abundant FADN data and a relatively high quality of this dataset the most suitable mixed-methods enabling the assessment of the RDP effect on the sectoral impact indicators should be driven by quantitative quasi-experimental approaches. Qualitative methods should be applied for triangulation of quantitative findings, and not as a replacement of quantitative methods. Qualitative methods may include interviews and focus groups with stakeholders, such as expert groups and focus groups, with a sample of farmers (beneficiaries and non-beneficiaries).[9] The added value of qualitative methods at this stage of the evaluation, i.e. when quantitative findings on programme net effects on competitiveness have been produced, consists of*:*

- The assessment of how important the quantitative findings at the local level are, i.e. how important is the heterogeneity of local conditions for challenging the worth of findings derived from standardised quantitative methods;
- The explanation of how and how the RDP effect has been achieved, e.g. clarifying the background mechanisms affecting the impacts (e.g. change of organisational behaviour, incentives, coordination, indirect 'trade-fair' effects that would not have materialised without the intervention).
- Explanation of the values, trends, patterns emerging from the quantitative analysis;
- Provide insight into how to design future policy approaches by interpreting the consequences of the quantitative findings.

Macro-level assessment

Access to data and its quality & creation of consistent database and data infrastructure

Statistical data showing developments of the three sector-related impact indicators are available from Eurostat at a country level (one value per calendar year). In Member States with regional RDPs no Eurostat information exists about the respective indicators for a given programming area. While

---

[8] See Guidelines: Assessment of RDP results: How to prepare for reporting on evaluation in 2017; HD, 2016.
[9] It is often possible to quantify qualitative information by Likert scores/discrete choice etc. and to, subsequently, estimate qualitative information on mechanisms with certain statistical procedures (e.g. linear and log-linear models for qualitative analysis, cf. Sanns and Schuchmann 2000, Lineare und loglineare Modelle in Psychologie und Sozialwissenschaften, München/Wien: Oldenbourg).

Eurostat impact indicators include information collected from *both* RDP supported- and non-supported farms, it would be very difficult for the evaluator to assess the net impact of the RDP on those indicators, without estimating relevant effects at the micro-level for both groups of farms. One of the possible alternative approaches that enables the calculation of RDP impacts at the macro- level could be an **application of a sectoral model** in which the respective sectoral impact indicators would be presented as respective model endogenous variables, containing explicit links to all individual effects originated from the RDP policies. However, the development of such a model (incl. data infrastructure), especially for the purpose of evaluation, is rather difficult and definitely would exceed time and budgetary constraints assigned to this specific evaluation exercise.

The macro level analysis (at the level of MA/regions) which is the ultimate objective of the impact evaluation in 2019 (i.e. the assessment of the sectoral effects represented by indicators I.01, I.02, I.03) is done via up-scaling of the already obtained micro level effects.

Selection of counterfactual option and macro-level method

The counterfactual analysis is done only at the micro level.

Yet, if an adequate **sectoral model** was available one could use a baseline scenario with 'all policies - in' as a baseline. This baseline would have to be confronted and compared with a specific policy scenario showing the development of the agricultural farms/sector *without the RDP* (counterfactual). However, in case such a model is not available, a reasonable solution is to define a counterfactual at a micro- farm level (see: description above) and extrapolate obtained results to all supported and non-supported farms, i.e. via bottom-up aggregation (or upscaling) defined at a programming area level.[10]

Net impact assessment at a macro-level

See above 'Net impact assessment' for micro-level and verification of obtained results through triangulation.

Micro-macro consistency and validation

If Approach A is applied, the consistency of evaluation findings concerning the programme's direct effects at micro and macro levels (programming area) is automatically warranted. In case of *indirect effects on supported and non-supported farms*, micro-economic findings after their aggregation can only roughly approximate the real scale of all possible indirect RDP effects. The main reason is a difficulty to explicitly model *all potential indirect* effects which non-supported farms could 'at least theoretically' have been confronted with.[11]

4.1.3 Approach B

Micro-level assessment

Access to data and its quality & creation of consistent database and data infrastructure

Like Approach A, we suggest using FADN for the assessment of the indicators' performance, I.01, I.02 and I.03. Some countries or regions might conduct a broader and more detailed farm survey than FADN,

---

[10] Unfortunately, there are no easily applicable techniques allowing for calculation of standard errors (level of significance),of results obtained at a micro-level. The problem is that the estimated variance of the treatment effect should also include the variation due to the estimation of the propensity score, due to the imputation of the common support, etc. A paper by Abadie and Imbens ('Matching on the estimated propensity score', Harvard University and National Bureau of Economic Research, 2012) establishes how to take into account the estimation of propensity scores. Sensitivity of obtained results can be addressed by applying the bounding approach proposed by Rosenbaum, 2002. Empirical application of the bounding approach can be found in Michalek, 2012.

[11] Validity of such generated results will inter alia depend on 'external' validity of PSM/DiD analysis. While applying PSM-DiD 'the common support region' can be used to make some judgements on the validity to transfer results from the analysed sample.

which can be used by evaluators. In each case it is important that data on eligibility criteria as well as on indicators are available in the chosen database. As pointed out in the section on Approach A, FADN data should satisfy the quality standards.

Selection of counterfactual option and micro-level method

The way the groups of supported and control farms are selected is best illustrated in Figure 1 below. In this example we consider that farms up to 90 hectares of utilised agricultural land are eligible for intervention. The farms in the band ±10 hectares around the cut-off point (90 hectares) are considered to be similar; the farms in the left band are supported (treated) and farms in the right band are the controls. Figure 1 shows the situation before (blue '*') and after (red '+') the intervention.

Figure 1.   Illustration of the Regression Discontinuity Design

Source: European Evaluation Helpdesk for Rural Development 2018

The following conditions are to be respected: [12]

- The cut-point is determined independently from the eligibility variable (that is, it is exogenous), and assignment to treatment is entirely based on the applicant ratings (related to the eligibility criteria) and the cut-point;

- Nothing other than treatment status is discontinuous in the analysis interval (that is, there are no other relevant variables in which observations on one side of the cut-point are treated differently from those on the other side);

- The functional form representing the relationship between the eligibility/rating variable and the outcome, which is included in the estimation model and can be represented by f(x) [x referring to eligibility/rating variable], is continuous throughout the analysis. [13]

Net impact assessment at micro level

The net effect is obtained as a comparison (difference) between values of the indicator in the bandwidth around the cut-off point of the eligibility of participation (see the box below). However, this effect is limited to the narrow group of beneficiaries around the cut-off point. In the simplest case, it is not clear to which extent the results are affected by other factors not considered in the linear regression. Controlling for the other factors influencing the indicator is possible but will make the approach more complex.

---

[12] See: 'A Practical Guide to Regression Discontinuity', Robin Jacob and Pei Zhu (2012),
https://www.mdrc.org/sites/default/files/RDD%20Guide_Full%20rev%202016_0.pdf

[13] This last condition applies only to parametric estimators.

Effects obtained by Regression Discontinuity Design

To model the effect of a particular measure of the RDP on individual outcomes y_i (impact indicator) through an RD approach, one needs a variable S_i that determines programme eligibility (such as size of holdings, etc. or scores given by a set of eligibility and preferential criteria) with an eligibility cut-off of s^* (e.g. €50 thousand or 50 points). The estimating linear regression equation is $y_i = \beta S_i + \varepsilon_i$, where individuals with $s_i \leq s^*$, for example, receive the programme, and individuals with $s_i > s*$ are not eligible to participate. Individuals in a narrow band above and below s^* are deemed to be 'comparable' since they would be expected to achieve similar outcomes prior to programme intervention.

If one assumes that limits exist on either side of the threshold s^*, the impact estimator for an arbitrarily small $\varepsilon > 0$ around the threshold would be the following:

$$E(y_i|s^* - \varepsilon_i) - E(y_i|s^* + \varepsilon_i) = E(\beta S_i|s^* - \varepsilon_i) - E(\beta S_i|s^* + \varepsilon_i)$$

Taking the limit of both sides of above equation as $\varepsilon \to 0$ would identify β as the ratio of the difference in outcomes of individuals just above and below the threshold, weighted by the difference in their realizations of S_i

$$\beta = \frac{y^- - y^+}{S^- - S^+}$$

Where the upper indices – and + indicate the limits from left and right.

Often in practice the determination or enforcement of eligibility is not 'sharp'. In this case, the discontinuity can be regarded -as stochastic or '"fuzzy,',"' meaning that it can be replaced with a probability of participating $P(S) = E(T|S)$, where $T = 1$ if support is received and $T = 0$ otherwise. Instead of measuring differences in outcomes above and below s^*, the impact estimator will measure the difference around a neighbourhood of s^*

Standard nonparametric regression can be used to estimate the treatment effect in either the sharp or the fuzzy regression discontinuity setup. For a sharp discontinuity design, the treatment effect can be estimated by a simple comparison of the mean outcomes of individuals to the left and the right of the threshold. Specifically, local linear regressions on the indicator y, given a set of covariates x, should be run for farms on both sides of the cut-off point, to estimate the difference

$$y^- - y^+ = \lim_{s_i \uparrow s^*} E(y_i|s_i = s^*) - \lim_{s_i \downarrow s^*} E(y_i|s_i = s^*)$$

It is recommended to proceed with the graphical analysis in the following steps:

Step 1: Make sure that the treatment is assigned exclusively on the basis of a cut-off value of the eligibility criteria.

Step 2: Regression discontinuity analysis should begin with a graphical presentation in which the value of the outcome (one of the indicators I.01, I.02, I.03) for each data point is plotted on the vertical axis, and the corresponding value of the eligibility variable/rating is plotted on the horizontal axis. The graphical presentation provides a powerful visualisation of discontinuity ('jump') in the outcome indicator at the cut-off point.

It is recommended to proceed with the graphical analysis in four steps:

- Divide the rating variable into a number of equal-sized intervals, which are often referred to as 'bins';
- Calculate the average value of the outcome variable and the midpoint value of the rating variable for each bin and count the number of observations in each bin;
- Plot the average outcome values for each bin on the Y-axis against the midpoint rating values for each bin on the X-axis, using the number of observations in each bin as the weight, so that the size of a plotted dot reflects the number of observations contained in that data point;
- To help readers better visualise whatever patterns exist in the data, one can superimpose flexible regression lines on top of the plotted data. This also provides a visual sense of the amount of noise in the data.

A challenge of the graphical assessment is to state bin width. There are some recommended formal tests which can be found in Jacob and Zhu (2012).

Step 3: Next, the evaluator can turn to estimating the treatment effects using a RD design formally. There are two types of strategies for correctly specifying the functional form between the eligibility/rating variable and the outcome indicator in a single-rating RD case:

- 'Discontinuity at the cut-point': This parametric strategy uses every observation in the sample to model the outcome as a function of the rating variable and treatment status. This method considers all available observations including those far from the cut-point score in order to estimate the average outcome for observations near the cut-point score. To minimise bias, different functional forms for the eligibility/rating variable — including the simplest linear form, quadratic, cubic, as well as its interactions with treatment — are tested by conducting F-tests on higher-order interaction terms and inspecting the residuals.

- 'Local randomisation': This nonparametric/local strategy adopts local randomisation for the estimation of treatment effects and limits the analysis to observations that lie within the bandwidth of the cut-point where the functional form is more likely to be close to linear. The main challenge here is selecting the right bandwidth. Once the bandwidth is selected, a linear [or polynomial] regression is estimated, using observations within one bandwidth on either side of the threshold.

Step 4: Assess the internal validity of RDD Impact Estimates. If the cut-point is to be chosen in the presence of knowledge about candidates' ratings, decision makers can locate the cut-point in a way that includes or excludes specific candidates. On the other hand, if ratings are determined in the presence of knowledge about the corresponding cut-point, they can be manipulated to include or exclude specific candidates. The methods that researchers can use to determine whether the ratings or cut-points could have been manipulated (that is, whether or not a RD discontinuity design is internally valid) include:

- Examination of the implementation process;
- Plotting the probability of receiving treatment as a function of the rating variable. For a valid RD design, there should be a discontinuity (or 'jump') at the cut-point in the probability of receiving treatment;
- Plotting the relationship between non-outcome variables and the rating variable. Non-outcome variables here refer mainly to potential covariates that, according to the theory of action, should not be affected by the treatment.

Step 5: Assess the precision of the estimates obtained from an RD design. This is something that is particularly relevant when using an existing data set. The precision of estimated treatment effects is typically expressed in terms of a minimum detectable effect (MDE) or a minimum detectable effect size (MDES). [14]

Step 6: Like Approach A, the results of RDD should be accompanied by a critical discussion of the obtained evidence including triangulation with other quantitative and qualitative findings.

Macro-level assessment

Also, Approach B holds that the macro level analysis of the sectoral effects represented by indicators I.01, I.02, I.03 is done as up-scaling of the already obtained micro level effects.

Net impact assessment at macro level

As pointed out earlier, the estimated impact only applies to the observations at, or close to, the cut-point. It would make the up-scaling very problematic, allowing for only a rough estimation of the sectoral effects. However, Lee and Lemieux (2010) argue that under certain assumptions of heterogeneity of observations RD design might have a lot in common with randomised experiments. Indeed, the resulting

---

[14] For more details and exact specification of the methods look at 'A Practical Guide to Regression Discontinuity', MDRC, pp. 100. https://www.mdrc.org/publication/practical-guide-regression-discontinuity, Jacob, R. Zhu, L., (2012).

cut-point population will comprise the full target population and generalisation of the assessment results and up-scaling can be made. The up-scaling and verification of results can be done in the same manner as for Approach A.

Micro and macro consistency and validation

Since the macro-level analysis is done via up-scaling of micro-data, the consistency check appears automatically.

Pros and cons of using the Approach A and B

Pros and cons for methods suggested for Approach A and B can be found in the Guidelines Ex post evaluation of 2013-2007 RDPs, Chapter 4.3.3.2.

Preconditions for applying Approach A and B

There are three issues to be considered in the application of approach A and B in the RDP evaluation: data availability, skills and technical equipment (software) available to the evaluator and the time for evaluation.

- Data availability: Is the MA aware of the availability and coverage of FADN or similar data sets (individual farm level data) for the analysis? Can these datasets be provided by the MA or are easily accessible for the evaluators? In the opposite case, it is not realistic to collect this data for the 2019 evaluation exercise, and simple quantitative (naïve) or qualitative approaches have to be used.
- Evaluator´s skills and technical equipment: Can the evaluator demonstrate sufficient skills and technical equipment in the application? In case of PSM/GPSM practical experience of previous working with the method is essential. In the case of RDD, the evidence of good econometric skill should be provided.

Time: Is there sufficient time allocated for the evaluation? Both approaches are time demanding if the assessment of the RDP impacts should be done in an appropriate way (at least 4-person months for the assessment of the sectoral impacts in case the PSM/GPSM is used and at least 2 people months for the use of RDD).

4.1.4 Adequateness of suggested evaluation approaches

The above described evaluation approaches are discussed below in terms of their capacity to fulfil the evaluation quality criteria: rigour, reliability, robustness, validity, transparency, credibility, practicability and cost effectiveness. Definitions of the quality criteria are provided in Table 2.

Table 2.   Evaluation quality criteria

Quality criteria	Definition
Rigour	Ability to produce exact findings. Rigorous evaluation requires first of all to be able to rely on a causal analysis. Rigour in causal attribution of the applied quantitative evaluation method (part of an overall evaluation design) comes very close to the ideal, i.e. experimental design.
Reliability	Quality of the collection of evaluation data when the protocol used makes it possible to produce similar information during repeated observations in identical conditions.
Robustness	Ability to produce findings which are stable and resilient to small but deliberate changes
Validity	Accuracy, logical and factual soundness of method in depicting the reality without errors and the conclusions and decisions based on this depiction.
Transparency	Transparency of an evaluation methodology requires that users know exactly its main elements, structure, parameters, rules and functional responses. A user can therefore monitor that they are followed. A valid estimate of the counterfactual should be based on clear and transparent assignment rules.
Credibility	Ability of the method to generate findings which can be trusted by stakeholders, for example the method demonstrates the causality, isolate programme effects from other factors, and eliminate the selection bias.
Practicability	Extent to which the method can be applied without adverse consequences (e.g. ethical) given the available data, resources, time.
Cost-effectiveness	Ability to provide sound evaluation findings whilst spending less money.

Table 3.   Adequateness of the proposed evaluation approaches for the assessment of CAP common impact indicators: I.01, I.02 and I.03

Quality criteria	Approach A (optimal) Micro-level **Combination of Propensity Score Matching methods with Difference in Differences Method (PSM-DiD)** Macro level **Upscaling macro-level findings** **Application of sector model**	Approach B (alternative) Micro-level **Regression Discontinuity Design (RDD)** Macro level **Upscaling macro-level findings**
Rigour	**Approach A is rigorous,** because it is based on well-developed theories, i.e. statistical theories about causal effects and concepts on how to measure these effects; it is widely applied and accepted in the scientific community; and it is described in methodological textbooks.	**Approach B is partly rigour.** The assignment near the cut-off can be seen almost as random. Further, individual units around the eligibility cut-off point (on both sides) might be regarded as similar; yet the selection bias in terms of other farms characteristics may be still large.
Reliability	**Approach A is highly reliable**. After collection of data it is possible to repeat all analytical steps and produce similar results during repeated observations in identical conditions.	**Approach B is reliable.** After collection of data it is possible to repeat all analytical steps and produce similar results during repeated observations in identical conditions.

Quality criteria	Approach A (optimal) Micro-level **Combination of Propensity Score Matching methods with Difference in Differences Method (PSM-DiD)** Macro level **Upscaling macro-level findings** **Application of sector model**	Approach B (alternative) Micro-level **Regression Discontinuity Design (RDD)** Macro level **Upscaling macro-level findings**
Robustness	**Approach A is partly robust**. Robustness of results shall be checked by applying sensitivity analyses, e.g. determining an influence of 'unobservable' on results obtained. Because of the assumption behind PSM that the applied characteristics (covariates) used to estimate the propensity score, explain all differences between the supported units and the comparison group prior to programme implementation, adding a new covariate might change counterfactuals and thus the estimated effects (ATT, ATE).	**Approach B is partly robust**. The choice of a bandwidth around a cut-off score is rather arbitrary and its change might affect results (captured effects).
Validity	**Approach A is partly valid.** Special statistical tests can be applied to judge the method's external validity enabling the transfer of results from the analysed sample to the whole population (i.e. via analysis of a 'common support region'). Validity depends on the number of units involved in the analysis. Generally, the larger the better. Up-scaling requires that the micro sample is representative. Information on this relation to the whole population is needed. The 'external validity' depends on the similarity between the populations where the sample results are to be transferred to.	**Approach B is partly valid.** The estimated impact is only valid in the neighbourhood around the eligibility cut-off score.
Transparency	**Approach A is transparent.** As the code of the analysis is written in a given programming language, users know exactly its main elements, structure, parameters, rules and functional responses. A follow-up is therefore relatively easy. However, its implementation highly depends on evaluators' skills.	**Approach B is transparent.** It allows to identify the programme's causal effect without imposing arbitrary assumptions on the selection process, functional forms, or distributional assumptions on errors. It can also marked as 'straightforward'.
Credibility	**Approach A is credible.** The applied method can be seen as a credible approximation of what would have occurred in the absence of the intervention, and to compare it with what actually happened. The method has been designed specially to enable isolation of programme effects from other factors, eliminate the selection bias and be conceptually straightforward (the idea of matching can be well communicated). This might increase its acceptability and credibility, but it highly depends on the evaluator´s skills.	**Approach B is partly credible.** The method can easily demonstrate the changes (effects) around the cut-off point. But the fact that the effects relate only to a narrow bandwidth around the eligibility threshold might be difficult to explain.
Practicability	**Approach A is partly practical.** The method is demanding in terms of data, time and skill (of the evaluator). FADN data are generally available for every Member State; however, their usefulness is contingent upon the link to other administrative data (identification of RDP beneficiaries). An	**Approach B is partly practical.** The method is relatively simple, but the estimated effects can hardly be extrapolated to population. The effects are not valid for all supported in the sample (the units distant from the cut/off).

Quality criteria	Approach A (optimal) Micro-level **Combination of Propensity Score Matching methods with Difference in Differences Method (PSM-DiD)** Macro level **Upscaling macro-level findings** **Application of sector model**	Approach B (alternative) Micro-level **Regression Discontinuity Design (RDD)** Macro level **Upscaling macro-level findings**
	important criterion for practicability is the necessary qualification of the expert carrying out the analysis applying this method. Also, time necessary for carrying out the analysis mainly depends on the experience of the evaluator with a given method. Regarding this criterion, the practicability of the method can be limited due to insufficient experts' knowledge and skills in this respect. On the other hand, the method´s practicability also lies in the fact that estimated effects hold for the sample and can be (if the sample is representative) extrapolated to the whole population.	
Cost-effectiveness	**Approach A is cost effective, but it depends on the data availability.** In case the data are not available from existing sources the method can be costly. In case the data are available (e.g. FADN-data), the approach is very well suited for large scale applications and can be very cost effective. Here, the key is to explore in advance (prior the evaluation) which data already exists and what still needs to be collected. Since FADN data is generally available and can be used in the form of a panel data, using them within a framework of counterfactual approach is always cost effective in comparison to alternative approaches, such as primary data collection and analysis via surveys. The approach can be completed with additional case study, costs remaining low compared to non-quantitative methods.	**Approach B is cost-effective,** since it is less demanding in terms of the amount of data needed for the assessment of the effects. This makes it cheaper comparing to PSM/DiD.

4.1.5 Examples of advanced evaluation methodologies (PSM/DiD, GPSM, coarsened exact matching, etc.) applied to Mid-term and Ex-post evaluations of RDP (incl. other studies):

- Austria: applied PSM/DiD methodology
  - o 'Investment support under Rural Development Policy'; final report of the DG-AGRI study; Metis, WIFO, Aeidl; November 2014; case study Austria, http://ec.europa.eu/agriculture/evaluation/rural-development-reports/investment-support-rdp-2014_en.htm

- Czech Republic: applied PSM/DiD methodology
  - o 'Investment support under Rural Development Policy'; final report of the DG-AGRI study; Metis, WIFO, Aeidl; November 2014; case study Cezch Republic, http://ec.europa.eu/agriculture/evaluation/rural-development-reports/investment-support-rdp-2014_en.htm

- Estonia: applied PSM/DiD methodology
  - o Ex Post Evaluation of the Estonian Rural Development Plan 2007–2013, final report (December 2016); https://www.agri.ee/sites/default/files/content/arengukavad/mak-2007/seire/erdp-2007-ex-post-evaluation-report.pdf

- France: applied PSM/DiD methodology
  - o Chabé-Ferret, S, and J Subervie (2013), How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DiD-matching, in: *Journal of Environmental Economics and Management, 2013*

    https://www.sciencedirect.com/science/article/pii/S0095069612000952
- Germany: applied PSM/DiD methodology
  - o Schleswig-Holstein: Michalek, J. (2016), 'Investment Crowding Out: Farm-level Evidence from Northern Germany', co-authored by P. Ciaian and d'A. Kancs, in: *Regional Studies*, vol. 50, issue 9*,* pp. 1579-1594, – DOI: 10.1080/00343404.2015.1044957. http://www.tandfonline.com/doi/full/10.1080/00343404.2015.1044957
  - o Schleswig-Holstein: 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, http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=5379
  - o Hessen: 'Investment support under Rural Development Policy'; final report of the DG-AGRI study; Metis, WIFO, Aeidl; November 2014; http://ec.europa.eu/agriculture/evaluation/rural-development-reports/investment-support-rdp-2014_en.htm
  - o Hessen: Pufahl, A., Weiss, C. R. (2009): Evaluating the effects of farm programs: Results from Propensity Score Matching. European Review of Agricultural Economics 36 (1), 79-101
  - o Bavaria: Michalek J., et.al. (2016), 'Feasibility Study on Modelling Rural Development Policies in IFM-CAP', Final Report, EC DG-JRC, Sevilla, October 2016.
- Latvia: applied PSM/DiD methodology
  - o Benga E. (2016); Economic impacts of semi-subsistence farm support measure of Latvian rural development programme 2007-2013; http://www2.llu.lv/research_conf/proceedings2016_vol_2/docs/LatviaResRuralDev_22nd_vol2-196-202.pdf
  - o Benga E., and J. Hazners, 2016; 'Counterfactual evaluation of farm modernisation measure of Latvian RDP 2007-2013 with Propensity Score Matching', https://enrd.ec.europa.eu/sites/enrd/files/gpw-03_3-1_latvia_hazners.rev_.pdf
- Poland: applied PSM-DiD, GPSM methodology
  - o 'Investment support under Rural Development Policy'; final report of the DG-AGRI study; Metis, WIFO, Aeidl; November 2014; case study Poland, http://ec.europa.eu/agriculture/evaluation/rural-development-reports/investment-support-rdp-2014_en.htm
  - o 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 http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=5360
- Romania: applied PSM/DiD methodology
  - o Ex-post evaluation study of the national rural development programme 2007-2013 (2017), final report: http://www.madr.ro/docs/dezvoltare-rurala/studii/Ex-post-evaluation-of-NRDP-2007-2013-EN.pdf
- Slovakia: applied PSM/DiD, GPSM methodology
  - o 'Ex post evaluation of Rural Development Programme 2007-2013 in the Slovak Republic'. Final Report, December 2016. http://mpsr.sk/index.php?navID=318&navID2=318&sID=43&id=12468 (English version of the report is available upon request)
  - o Michalek, J., P. Ciaian and J. Pokryvcak (2018), 'The impact of Producer Organizations on farm performance: The case study of large farms from Slovakia', in: Food Policy 75 (2018), 80-92; https://www.journals.elsevier.com/food-policy
  - o 'Investment support under Rural Development Policy'; final report of the DG-AGRI study; Metis, WIFO, Aeidl; November 2014; case study Slovak Republic, http://ec.europa.eu/agriculture/evaluation/rural-development-reports/investment-support-rdp-2014_en.htm

- o 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, http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=5379
- o 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 http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=5360
- 15 EU Member States: applied GPSM methodology
  - o Michalek J., P. Ciaian and d'A. Kancs (2014), 'Capitalization of CAP Single Payment Scheme into Land Value: Generalized Propensity Score Evidence from the EU', in: *Land Economics*, University of Wisconsin Press, May 2014 vol. 90, No 2, pp:260-289. http://le.uwpress.org/content/90/2/260.full.pdf+html
- Other non-EU countries: use of survey data for application of PSM/DiD methodology
  - o Kosovo: 'Impact and In-depth Study of the Rural Development Grant Programme in Kosovo', Final Report, ADE, March 2016.

4.1.6 Complementary guidance for calculation of I.01, I.02 and I.03

The guidance below complements the guidance for calculation of I.01, I.02 and I.03 at the micro-level as provided in the Chapter 2.2 of PART II of the Guidelines and Fiches for impact indicators I.01, I.02 and I.03. The fiches contain the definition of the indicators, the methodology how to obtain the macro-level data for the numerator, denominator or directly the index of the measurement unit, as well as the data sources. The following table summarises the methodology for the assessment of the indicator at the micro-level (individual farm) while using the FADN as data source. The micro-level assessment is important to net out the macro- level value of impact indicator. **This guidance is not compulsory.**

INDICATOR	I.01
Name of the indicator	**Agricultural entrepreneurial income (AEI)**
Description of indicator	See the fiche for common CAP indicator I.01
Calculation of indicator	Calculation of I.01 should be based on the FADN variables: SE135 = Total Output crops and crop production SE206 = Total Output livestock and livestock products SE275 = Total intermediate consumption SE360 = Depreciation SE600 = Balance current subsidies and taxes SE365 = Total external factors (wages, rents and interest paid) SE010 = Total labour input in full time equivalents The formula is as follows: (SE135 + SE206 − SE275 −SE360 + SE600 − SE365)/SE010 Where: AEI = agricultural entrepreneurial income calculated at micro-level
Data source	FADN (individual data on the request)
References/location of the data	FADN – national liaison offices
Data collection level	Farms
Frequency	Annual data available
Delay	Y-2 (two years)
INDICATOR	I.02
Name of the indicator	**Agricultural factor income (AFI)**
Description of indicator	See the fiche for common CAP indicator I.01
Calculation of indicator	Calculation of I.02 should be based on the FADN variables: SE135 = Total Output crops and crop production SE206 = Total Output livestock and livestock products SE275 = Total intermediate consumption SE360 = Depreciation SE600 = Balance current subsidies and taxes SE010 = Total labour input in full time equivalents The formula is as follows: (SE135 + SE206 − SE275 −SE360 + SE600)/SE010 Where: I.02 = agricultural factor income calculated at micro-level
Data source	FADN (individual data on the request)
References/location of the data	FADN – national liaison offices
Data collection level	Farms
Frequency	Annual data available, calculations on request (FADN)
Delay	Y-2 (two years)

INDICATOR	I.03
Name of the indicator	**Total factor productivity in agriculture (TFP)**
Description of indicator	See the fiche for common CAP indicator I.01
Data source and calculation of indicator	The main source of data will be Farm Data Accountancy Network (FADN) in the Member State conducting the evaluation, Eurostat/national/regional statistical office (for prices and wages) and additional national sources which might allow deeper detail or substitute for missing information.

The main source of data will be Farm Data Accountancy Network (FADN) in the Member State conducting the evaluation, Eurostat/national/regional statistical office (for prices and wages) and additional national sources which might allow deeper detail or substitute for missing information.

We suggest using the following data:

Output (n=3): Crop Production (FADN SE135), Livestock Production (FADN SE206) and Other Output (SE256) in nominal (basic) values.

Inputs/Factors (m=4): Labour in AWU (FADN SE010), UAA (FADN SE025) in hectares, Working Capital (FADN SE275 [intermediate consumption]) in nominal value, Fixed Capital (FADN SE360 [depreciation]) in nominal value.

By deflating nominal values by price indices referring to the base period we yield real values which represent quantities (q_i^t and x_j^t).

The input and output price indices cannot be found in FADN or derived from FADN data. Beside national statistics, suitable might be the price statistics of Eurostat:

Price indices of agricultural products, output (2010 = 100) - annual data [apri_pI.10_outa].

Crop output price index, including fruits and vegetables (code 1000000).

Animal output price index (code 130000).

Agricultural output price index (140000) can be used in reference to other output.

Labour cost data (per AWU or an index over time) can be derived from the Economic Accounts for Agriculture (by dividing *Compensation of employees* [aact_eaa01] by *Salaried labour input in AWU* [aact_aII01]) or from the Labour Costs Statistics of the Eurostat.

The land prices (indices) can be obtained from Eurostat statistics 'Land prices and rents - annual data [apri_ap_aland]' from the Economic Accounts for Agriculture

Price indices of the means of agricultural production – inputs (w_j^t) (2010 = 100) - annual data [apri_pI.10_ina].

Goods and services currently consumed in agriculture (code 200000) – applied to Working Capital ($w_{working\ capital}^t$).

Goods and services contributing to agricultural investment (code 210000) – can be applied to Fixed Capital, but with some reservation (and further adjustment) since the depreciated capital (as the whole fixed capital) is formed by investment of various ages. If there is knowledge of the details the index can be adjusted ($w_{investment\ good}$).

In addition, fixed capital carries cost of resources used for its purchase (investment) – interest. It can be assumed that investment resources have opportunity cost and thus the replacement value of the consumed capital is multiplied by interest rate corresponding to long term government bond yield.

$$w_{fixed\ capital}^t = w_{invest.good}^t * interest\ rate$$

at time t.

Because most of the available figures of the output aggregates and the factors from the FADN will be in nominal values of production and factors, the I.03 indices can be straight forward calculated by multiplying or dividing the monetary values by price ratios $\frac{p_i^t}{p_i^0}$, $\frac{w_j^t}{w_j^0}$ (indices).

$$TFP_L^s = \frac{Q_L^s}{X_L^s} = \frac{\left. \sum_{i=1}^{n} \frac{p_i^0}{p_i^s} * r_i^s \middle/ \sum_{i=1}^{n} r_i^0 \right.}{\left. \sum_{j=1}^{m} \frac{w_j^0}{w_j^t} * c_j^s \middle/ \sum_{j=1}^{m} c_j^0 \right.}$$

INDICATOR	I.03
Name of the indicator	Total factor productivity in agriculture (TFP)
	$$TFP_L^s = \frac{Q_L^s}{X_L^s} = \left[\frac{\sum_{i=1}^{n} \frac{p_i^s}{p_i^0} * r_i^0 \Big/ \sum_{i=1}^{n} r_i^s}{\sum_{j=1}^{m} \frac{w_j^s}{w_j^0} * c_j^0 \Big/ \sum_{j=1}^{m} c_j^s} \right]^{-1}$$ Where $r^t = (r_1^t, ..., r_n^t)'$ and $c^t = (c_1^t, ..., c_m^t)'$, are vectors of revenue and cost shares respectively; t=0 and s. Note that labour cost and land rental cost are to be calculated for all used factors (hired and own labour and hired and own land).
References/location of the data	FADN – national liaison offices Eurostat web portal: http://ec.europa.eu/eurostat/data/database National statistical offices
Data collection level	Farm level
Frequency	Annual data available
Delay	Y-2 (two years)
Comments/caveats	The level of detailed information required to compile the indices makes the exercise demanding in terms of time. The calculation of the TFP index is complex and might lead to mistakes. When using price indices from statistics, please make sure that they refer (recalculate them) to the same base year.

4.1.7 Dos and don'ts

Dos

- Use farm book-keeping data in order to calculate an equivalent of I.01, I.02 and I.03 indicators (or additional indicators) at the micro-farm level in case of delays in FADN.
- Estimate first the net direct effects and second the net indirect effects of the RDP on the three sectoral impact indicators. Use micro-data (separately on supported and non-supported farms) and aggregate them to the macro-level.

Don'ts

- Calculate net effects of the RDP by using the sectoral impact indicators as outcome, calculated by Eurostat at the Member State level.
- Present the effects calculated by RDD as valid for the whole population.

4.2   CAP common impact indicator I.07

The following sections of the Technical Annex are related to Chapter 2.3 in PART II of the Guidelines.

4.2.1 Additional indicators (examples)

Additional indicators aim to support the evaluator in putting GHG emissions and ammonia emissions from agriculture into a wider frame and perspective. In this context, additional indicators do not substitute impact indicators. Indicators on GHG emissions from livestock and/or managed soils and ammonia emissions disaggregate emissions into their sources and show if the RDP has targeted the correct 'emission sources'. The status of manure storage and of tillage practices show if there is a need to target and further support these activities that reduce GHG and ammonia emissions. Livestock trends will indicate whether agriculture will have to face increasing or decreasing emissions in the future. For example, with regard to ammonia emissions, knowing the quantity and type of livestock is not sufficient because the amount emitted by livestock can be a function of many variables. These include properties of the animal manure which in turn depend on the animal feed, its age and weight. These factors affect the efficiency of the conversion of nitrogen in feed to livestock production (milk, eggs, etc.) and hence the remaining nitrogen in the manure and the proportion of that nitrogen that is volatised. Additionally, how manure is managed (liquid or litter) and how manure is stored (open or covered tanks) is very important. Soil properties affect the proportion of nitrogen converted into ammonium and then into nitrate and this determines the amount emitted. Not all, but some of these indicators may be extracted from the SAPM and be useful in showing a baseline (even if it is dated) in anticipation of a replication of the SAPM in the future. Evaluators may want to show the contributions of the RDP to reducing emissions from fuel combustion, by highlighting the RDP's activities of energy substitution (energy efficiency) and their numerical contribution without counting them in the indicator I.07. Chapter CRF1A4C on 'Fuel combustion in agriculture, forestry and fishing' provides a good estimate of the contribution of the primary sector[15], not of agriculture alone. Finally, LULUCF contributions may be further exemplified if there are data in the relevant Chapters. Suggested additional indicators are briefly presented in Table 4 below.

---

[15] See Eurostat data at: http://ec.europa.eu/eurostat/en/web/products-datasets/-/env_air_gge .

Table 4.   Examples of additional indicators related to emissions from agriculture

Indicator	Measurement Unit	Use of additional indicator	Data sources and frequency of collection
GHG from Livestock	1) Share (%) of GHG from livestock to total agriculture emissions.	Reveals if the livestock sector is an important issue for GHG emissions and its trend overtime.	Eurostat variable 'env_air_gge' that is reported yearly with 2 years lag.
GHG from Managed Soils	1) Share (%) of GHG from managed soils.	This indicator captures human induced additions of N to managed lands. This refers especially to synthetic and organic N fertilisers and urea, amongst others.	Eurostat variable 'env_air_gge' that is reported yearly with 2 years lag. Include both direct and indirect emissions.
Ammonia emissions	1) Total ammonia emissions in tonnes. 2) Share (%) of agricultural ammonia emissions.	The indicator I.07-2 only measures the absolute emissions from the agricultural sector and does not relate to the total ammonia emissions to reveal if agriculture is a significant contributor and its trend overtime.	Eurostat variable 'env_air_emis' that is reported yearly with 2 years lag.
Manure storage	1) Share (%) of holdings with livestock which have manure storage facilities in total holdings with livestock. 2) Share (%) of manure applied with different techniques and manure incorporation time.	Manure storage is AEI 11.3[16] and is related to GHG emissions and ammonia emissions. It shows the extent of the issue.	Survey on agricultural production methods (SAPM) has information related to 2010.[17] It is not known whether the SAPM will be repeated and if data will be replicated.
Livestock trends	1) Number of animals of cattle, equidae, sheep, pigs and poultry in LSU. 2) Share (%) of major livestock types (cattle, equidae, sheep, pigs and poultry) in total livestock population.	Livestock is the main contributor of emissions to agriculture, however, not all livestock have the same contribution. Livestock patterns are portrayed by AEI 10.2.[18]	The FSS has data[19] to the NUTS 2 level up to 2013 (latest survey) for all EU Member States. The 2016 survey is still ongoing in certain Member States while others have already published provisional results.
Tillage practices	1) Share (%) of arable areas under conventional, conservation and zero tillage. 2) Arable areas under convention, conservation and zero tillage in '000 ha.	Tillage practices describe the share of arable areas under conventional, conservation and zero tillage and is AEI 11.2.[20]	Survey on agricultural production methods (SAPM) has information related to 2010[21]. It is not known whether the SAPM will be repeated and if data will be replicated.

---

[16] AEI 11.3 is accessible at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_manure_storage
[17] Eurostat data at the NUTS 3 level accessible at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=aei_fm_ms&lang=en
[18] AEI 10.2 is accessible at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_livestock_patterns
[19] Eurostat data accessible at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=aei_ef_ls&lang=en
[20] The AEI 11.2 is accessible at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_tillage_practices
[21] Eurostat data are accessible at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Farm_structure_survey_2013_-_main_results

4.2.2 Approach A

Micro-level assessment

Access to data and its quality & creation of consistent databases and data infrastructure

GHG emissions are reported in tonnes of CO_2 equivalents and ammonia emissions in Kilotons or tonnes of NH_3. The time series for both indicators I.07-1 and I.07-2 are readily available at national level with a lag of almost two and a half years.[22] An issue may arise when the indicators need to be calculated at regional level. The EU reports annual GHG inventories to the United Nations under the United Nations Framework Convention on Climate Change (UNFCCC). The national submissions are considered part of the EU inventory and contain information about the methodology used to estimate emissions.[23] Following the reports, and especially the data sources with the adopted coefficients, the evaluator will be able to regionalise both indicators.[24]

> Eurostat's data for estimating the indicators at national level are two years behind. This implies that the latest data for assessment of impacts in 2019 will be those of 2016. However, the national authority may provide estimates for 2017. As concerns regional data, few Member States provide regional data for GHG emissions. The regional disaggregation of such data also varies among Member States: some provide a very low level of spatial disaggregation while others provide (upon request) a very high disaggregation (up to the NUTS 3 level).
>
> If there is no access to regional data there are two options:
>
> a) 'coefficients' (for various types of livestock, farmland, etc) which mechanically produce GHG emissions based on regional stock and activities. In this case estimates are calculated easily taking into consideration that there is a consistent series of stock (livestock) and activities for the regions, to which you apply the tier 2 coefficients.
>
> b) 'approximating' the regional GHG emission e.g. via running a very simple time-series analysis of national GHG emission on the basic components of emissions i.e., livestock numbers by category, cropland by category and grassland by category to examine how these contribute to the overall formulation of national GHG emissions. Then using these regression coefficients and regional data to produce a regional 'approximation' to GHG emissions.
>
> A good national time series (per activity and per type of gas) can be extracted directly from UNFCC at: http://di.unfccc.int/ghg_profile_annex1

At micro level, the information required to estimate deadweight does not exist in readily available databases. Thus, the evaluator will have to collect it through a survey of RDP supported and non-supported agricultural holdings as there are not any other ways to retrieve or simulate this information.

---

[22] Data for all I.07 indicators are available at national level from Eurostat at: http://ec.europa.eu/eurostat/en/web/products-datasets/-/env_air_gge

[23] Not all Member States use the same methodology. The detailed methodology and data sources for calculating various 'emissions' is presented in Annex III of the latest 'Annual European Union greenhouse gas inventory 1990–2015 and inventory report 2017' accessible at: https://www.eea.europa.eu/publications/european-union-greenhouse-gas-inventory-2018 . The latest national inventory database is accessible at: https://www.eea.europa.eu/data-and-maps/data/national-emissions-reported-to-the-unfccc-and-to-the-eu-greenhouse-gas-monitoring-mechanism-14/national-greenhouse-gas-inventories-ipcc-common-reporting-format-sector-classification
National inventory submissions are also accessible via the UN's UNFCCC portal at:
http://unfccc.int/national_reports/annex_i_ghg_inventories/national_inventories_submissions/items/8812.php

[24] Further assistance on methodology is provided by Eurostat's manuals 'Analysis of methodologies for calculating greenhouse gas and ammonia emissions and nutrient balances' accessible at:
http://ec.europa.eu/eurostat/documents/3888793/5850629/KS-RA-11-024-EN.PDF and 'Farm data needed for agri-environmental reporting' accessible at: http://ec.europa.eu/eurostat/documents/3888793/5849721/KS-RA-11-005-EN.PDF

Selection of a counterfactual option and micro-level method

In netting out the effects of RDP on GHG emissions and ammonia the evaluator should be aware that the RDP measures primarily linked to FA5D exercise a direct, presumably positive, effect while measures linked to other FAs, and especially FAs of priority 4, exercise an indirect, positive or negative, effect. Non-supported agricultural holdings are affected indirectly by RDP. Thus, the two primary comparison groups are RDP supported and RDP non-supported agricultural holdings. If the RDP targets GHG emissions and Ammonia reductions through different and well populated measures, then the evaluator may consider to sub-divide the group of supported farm holdings to sub-groups according to the implemented measures.

The micro-level assessment should follow these distinct steps:

Step 1: *Recognise the institutional framework* within which GHG emissions from agriculture are measured (methods) and reported to the EU.

Step 2: *Calculate I.07 indicators* if they are not readily available from Eurostat (case of regional RDPs). This step is depicted by boxes numbered 1 and 2 on the upper part of Figure 7 in Chapter 2.3.1 of PART II. In the case of regional RDPs in which regional GHG and ammonia emission data cannot be produced, the evaluator can approximate a GHG emission indicator by adopting a statistical methodology. For example, this methodology may be a time series regression of national data on key variables that formulate the emissions level such as livestock number by type, cropland by type, grassland, etc. This will not be an 'exact' measurement of the emissions but it will establish a baseline and the consequent changes. Even this will provide a broad measurement of the RDP's gross effects.

Step 3: *Retrieve Result Indicators R16, R17, R18, R19* that will reveal the size of the supported agricultural holding population and the variety of measures used within the RDP's intervention logic. Retrieve IACS/LPIS data to examine the spatial coverage of the supported agricultural holdings. Retrieve soil and land use maps to examine the degree of environmental heterogeneity in the areas covered by the RDP's measures. This step is depicted by the two boxes numbered 3 and 4 in the middle of the above-mentioned Figure 7.

Step 4: Decide if the number of supported agricultural holdings (from Step 3 above) is sufficient for carrying out a survey, i.e. if there is scope in establishing comparison groups. If the evaluator decides not to carry out a survey, the flow of operations from the box number of Figure 7 named 'comparison groups exist' will follow the 'No' branch leading to the Approach in 2019.

Step 5: *Set up the survey:*

a)   Comparison groups will be created and the flow of operations from box number 5, of above mentioned Figure 7, will lead to box Number 7, the Approach A.

b)   Decide on the group or sub-groups of supported holdings (from Step 3 above) and the counterfactual, i.e. holdings that did not receive support from the RDP.

c)   Decide on the spatial coverage of the survey as informed by Step 3 above. Decide on sample sizes depending on the method to be employed, the cost of sampling and the available budget.

Step 6: *Design a questionnaire* that will capture GHG and ammonia emissions, and GHG and ammonia emission changes on the holding. This work may be assisted by on farm GHG and ammonia emission calculators.

Net impact assessment at the micro-level: Once the data has been collected, coded and stored in a database, it can be analysed with an aim to estimate the net effect, scale up the estimations to the RDP level and verify the produced estimations with other sources by proceeding with the following steps.

Step 7: *Apply a method for analysing the data.* Two non-experimental (simple regressions and instrumental variables) and one quasi-experimental (matching) methods are suggested: Simple

regression on GHG emission and/or ammonia emissions with carefully chosen control variables that will reduce (but never eliminate) selection bias, estimated by a single or two stage process:

- Employ Instrumental Variables (IV) analysis that deal better with selectivity but are more demanding econometrically;
- Construct a matching counterfactual from the sample of non-supported farm holdings with a matching algorithm.

In the case where the RDP has access to its own regular survey of farm holdings and can create a panel of data, the above described methods can be coupled with Difference in Differences methods.

Step 8: *Estimate the Average Treatment Effect on the Treated* (ATT) and compute the RDP's net direct effect on reductions of GHG and ammonia emissions. The ATT is an estimate of the difference of what actually happened due to the intervention minus what would have happened in the absence of the intervention. In other words, a measure of the intervention's net effect.

Step 9: If there are indications of *important indirect effects, either on supported or non-supported agricultural holdings due to GHG emission* and ammonia reduction measures, which is rather unlike, these should be treated separately. Qualitative methods and especially in-depth interviews with experts will allow the evaluator to highlight at least these possible effects. The questionnaire also record transverse effects.

Step 10: *Aggregate the results and estimate the effects of the RDP at macro level.* At this stage the evaluator has two options. If the survey is representative of the RDP territory, the evaluator can scale up the micro results for both supported and non-supported farm holdings. In the case where the survey is not representative of the whole territory but only of specific targeted areas, the evaluator can work differently. One solution is to post-stratify the survey and assign weights to the cases to make them representative of the entire population. This is rather cumbersome and data demanding. Another solution is to apply the net direct effects coefficients to the total RDP effects as these are captured by the relevant Results Indicators and estimate only the total RDP net direct effect and not the total agricultural emissions.

Step 11: *Verify the results obtained* by the afore mentioned process with qualitative data obtained from interviewing experts, published case studies carried out in the RDP territory or in other RDPs facing similar agricultural conditions.

Macro-level assessment

Access to data and its quality and creation of consistent databases and data infrastructure

Data on GHG emissions from agriculture at NUTS 3 (or lower) level usually exist either in environmental databases or as an input in calculating national GHG emissions. These data, if they do not exist or if the evaluator cannot access it, it can be estimated using the national methodology for estimating GHG emissions, (tier 1 or tier 2) taking time constraints into account. Use of spatially disaggregate data must be evaluated on a case-by-case basis. In addition, it is important that the evaluator examines whether spatial data can be coupled with a national or European environmental database at the same spatial level. If there are no suitable data then, a macro-level assessment cannot be carried out.

Selection of a counterfactual option, macro-level method

Depending on data availability, the number of spatial units (NUTS 3 or smaller) and associated environmental and monitoring data, the evaluator can consider two alternative options. One is to construct the counterfactual from areas that do not get any RDP support. However, as this is rather rare, the evaluator can use a Generalised Propensity Scoring Matching (GPSM) algorithm for setting up the counterfactual. In this approach, as counterfactual are considered the areas below a 'threshold' of support.

Net impact assessment at a macro-level

The Generalised Propensity Scoring Matching (GPSM) can evaluate deadweight and produce net RDP effects in a macro assessment framework.

Micro-macro consistency and validation

For the net impact assessment, it is important that the results of both the micro and the macro-level assessment are consistent, i.e., results of these assessments show the same trend in relation to impact, even though the evaluator used slightly different methods for the assessments. Triangulation of net impacts with information from qualitative sources, published case studies or academic work should be performed. (see PART II, Chapter 2.1).

Pros and cons of using the Approach A

Important feature	Pros	Cons
Simple regression methods do not demand high econometric skills but their ability to reduce selection bias and produce unbiased net effects estimates depend on careful design of the survey and collection of 'good' control variables.	Easiness of application.	Possibility of bias is not eliminated.
Instrumental Variables (IV) approach is econometrically more demanding than a simple regression but, if there are suitable instruments, it deals well with selection bias.	Selection bias is adequately dealt with.	Intermediate econometric skills are required.
Matching algorithms and especially the Propensity Score Matching, is the best method for dealing with selectivity bias but it requires larger sample size for non-RDP participants in order to ensure that the largest proportion of RDP participants can be matched to a non-participant.	Best method for addressing selection bias.	Larger sample sizes and intermediate econometric skills are required.

Preconditions for applying Approach A

> If there is a sufficient number of supported agriculture holdings and the MA decides to use the survey in the assessment, then it is important to examine the possibility to conduct the combined survey to collect data & information on all environmental indicators (e.g., water abstraction, GNB, soil erosion). This will assign scale economies in sampling and can reduce the sampling cost dramatically. The following issues shall be clarified for the applicant of the survey in the RDP evaluation:
>
> - What will be the sampling procedure and sampling size based on your knowledge of the measures' uptake?
> - Will it be an RDP territory-wide survey or will target specific RDP 'problematic' areas? Will it address a wide range of measures or will be restricted to only a few (2-3) measures?
> - What will be the survey methods to collect data & information? e.g. telephone, email, personal interviews?
> - From which sampling frame are you going to choose non-RDP participants and how will you motivate them to respond to your survey?
> - Will you provide and/or use monitoring and application data for RDP participants in order to reduce the length of the interview by pre-filling some of the required information?
> - Will you use a GHG and ammonia emissions calculator for estimating if GHG and ammonia emissions change uniformly with one method?
> - Will you require specific econometric skills and proven experience from those who will analyse the survey data?
> - Will you require a specific 'upscaling' procedure based on the survey results?

4.2.3 Approach B

Micro-level assessment

We do not propose a specific micro-level assessment. One could propose simulation models on a typical landscape of the RDP, but these models are complicated and, in our opinion, outside the scope of an RDP evaluation exercise. However, if local universities or research institutes have developed and calibrated such models within the RDP's territory, evaluators could use it for the RDP evaluation.

Macro-level assessment - Naïve Baseline Comparisons

The Approach B is used in case there is no comparison group and the evaluator cannot set one up, and/or in the situation where a sufficiently accurate model[25] does not exist (box 6). Assuming there will not be time for evaluation, this directly takes the suggested methodology to a qualitative assessment of the RDP's net direct effects (see points 5 – 9 of the Figure 7 in Chapter 2.3.2 of PART II). Useful sources for framing the behaviour of non-supported RDP livestock and farm holdings may be found by instructing livestock experts and GHG and ammonia emissions experts in agriculture. This information can be useful in deducing the general trend prevailing over non-assisted livestock and farm holdings. This can provide an indication (not a precise estimate) of deadweight loss but does not give indications of leverage or other secondary effects. Finally, information may be collected for the role of supporting measures, such as M01 and M02 or M16.5 and also about the region's contribution to adopting (or even generating) innovative approaches. This information may be collected by instructing advisors and

---

[25] Simulation models, for the emissions of ammonia and nitrous oxide, do exist but they are complicated, and we hesitate to suggest them in the framework of an evaluation. A very good comparative study of various simulation models can be found in 'Comparison of land nitrogen budgets for European agriculture by various modelling approaches', Environmental Pollution, 159, pp.3254-3268, de Vries et al., (2011).

training authorities or by instructing possible innovation and cooperation partners to flag possible leverage and transverse effects. The contact can take the form of a series of in-depth interviews, a focus group or a Delphi evaluation or a MAPP exercise (see the introduction to environmental indicators for using the MAPP method in the assessment of environmental indicators).

As described in the text the Approach B is a qualitative approach that aims to approximate a 'net effect coefficient – deadweight coefficient', in the absence of the proposed farm holding survey or other detailed quantitative data. This means that, at best, once the chosen qualitative exercise is finished, the evaluator will be able to come up with a statement like 'the measure X aiming to reduce GHG emissions from enteric fermentation is estimated to have a deadweight of 80% (or a range, e.g., 60-70%) because the experts and stakeholders involved in the methodology (focus group, Delphi, etc.) provided evidence and agreed that the measure has already been adopted by non RDP supported farms to an extent equivalent to 20% of RDP supported farm holdings. If the monitoring data can show what the RDPs gross results were, (from complementary indicators R18 and R19) it is possible to apply this coefficient to estimate the RDP's net effects for deadweight. In the example above, it will be 80% of the part of R18 that targeted enteric fermentation by using measure X. In case there are no figures for R18 and R19, it is still possible to analyse a few typical projects to see what the average correspondence is between R18 and R16, and R19 and R17 to get an approximation of R18 and R19 indicators.

Micro-macro consistency and validation

Since there is no micro approach, there is no scope for a micro-macro consistency check. For a triangulation it is important to compare the findings from the qualitative approach with published reports and academic work on the effects of the measures, in the same or other RDPs, in a relevant socio-economic and physical context.

Pros and cons of using the Approach B - MAPP method

Important feature	Pros	Cons
General.	MAPP is particularly suited for analysing of more complex and long-term objectives, that cannot usually be evaluated with the help of one or more quantitative indicators.	
Implementation level: ideally local/regional.	At a lower level (local) it is easier to include in the focus group beneficiaries and non-beneficiaries. Focus group participants at local/regional level are more aware of the development trends in their area due to closeness. They are also more aware of the impact of other programmes.	It may be expensive to have several focus groups if the RDP covers a large region (e.g. Andalucía in Spain) or a whole country (e.g. Greece). In this case, only a representative number of areas/regions should be selected, based on clear criteria.
MAPP can assess quite a number of measures at the same time.	The method allows to focus on the most important measures in terms of funding.	None
MAPP can assess several impact and result indicators at the same time.	MAPP allows to assess the evolution of indicators which are hard to quantify, based on a point system and a descriptive assessment. Its systematic approach and the use of a point system produce results of greater external validity than purely qualitative data, e.g. derived from interviews or traditional focus group discussions.	MAPP does not give a quantitative value to indicators. MAPP is a qualitative tool with some quantitative elements. Though it uses a point system, its results cannot be used for statistical analysis.
MAPP allows the qualitative assessment of net impacts.	With MAPP, a specific programme is evaluated in relation to other ongoing programmes and/or other external factors.	It cannot replace robust quantitative methods.

Important feature	Pros	Cons
	Thus, net impacts can be estimated against gross development trends. It helps to bridge the 'attribution gap', i.e. the gap between outcomes that can directly be attributed to a specific programme and outcomes that are also influenced by other factors. It has an open context-orientated approach that allows the identification of not only planned, but also unplanned impacts.	It cannot quantify planned or unplanned impacts.
MAPP is flexible in the choice of participants and non-participants in the intervention, in line with the purpose of the assessment.	a) If MAPP is implemented at local level, beneficiaries and non-beneficiaries should be chosen b) If MAPP is implemented at regional/national level, representatives of beneficiaries and non-beneficiaries (e.g. farmers' association) should be chosen.	MAPP is an interactive method and as such its results depend strongly on the commitment of participants and their interest in the discussion.
MAPP offers a range of tools.	Pre-defined tools allow for a more structured and focused discussion. The tools can convert perceptions and experiences to quantitative information through the point system.	Despite the point system in the tools, this is not a quantitative method and should not replace such methods.
MAPP allows the assessment of deadweight.	Can be assessed with the point system.	Cannot give a numeric value of deadweight like quantitative methods.
MAPP allows to assess the causal links.	MAPP can assess the causal links between the relevant RDP measures and the effects on the environment.	It cannot quantify these links, only estimate their size or intensity.

Pros and cons of using the Approach B - Delphi method

Important feature	Pros	Cons
General.	The method allows experts to comment on their own forecasts and on the answers of others, in research of some consensus without the pressure of the group leader.	More pertinent for social analyses/evaluations. For environmental indicators, the information is more accurate/ clear-cut and several rounds of survey may not change the opinion of the expert.
Participants 'collective intelligence' feature.	Allows for a dialogue between geographically dispersed panel experts. Can take place virtually or face to face. Structured/organised group communication process. Flexible in terms of participants (from a few to hundreds, although frequent size is 20).	Cannot guarantee the commitment of experts in each round, especially if carried out virtually. There are no guidelines for determining sample size and sampling techniques. Requires a very competent panel facilitator. Limited group dynamics when done virtually.
Survey conducted in several rounds.	A key feature of the method is that the evaluation question can be answered in several rounds, to reach consensus on the 'right' answer.	It is more pertinent for forecasts than identifying 'what has happened'.
Timescale - limited amount of time required to complete the surveys.	Useful when there is no time or resources to organise face-to-face meetings/focus group.	However, each round of the survey requires time for data collection (multiple data collection, analysis, processing). If done virtually, requires time and participant commitment. Requires a lot of time to conduct several rounds.
Analysis of results.	Can condense experts' opinions in a few precise and clearly defined statements.	There are no guidelines for determining consensus. Complex data analysis.

Preconditions for applying Approach B

The key issues to be considered if using qualitative methods in evaluation are:

- Will the range for an approximate 'net effects' coefficient be quantified through the qualitative method?
- How are participants (experts and stakeholders) to be chosen?
- How many meetings, either consecutive of the same group or regional ones with different groups, will you set up?
- Will specific skills and proven experience in qualitative environmental assessments be required from those who will carry out the qualitative method?

4.2.4 Adequateness of suggested evaluation approaches

The above described evaluation approaches are discussed in the following table as regards their adequateness in fulfilling the evaluation quality criteria: rigour, reliability, robustness, validity, transparency, credibility, practicability and cost effectiveness. Definitions of quality criteria are provided in Table 2.

Table 5. **Adequateness of the proposed evaluation approaches for the assessment of CAP common impact indicator I.07 – GHG emissions from agriculture**

Quality criteria	Approach A (optimal)	Approach B (alternative)
	Micro-level **Regression and matching techniques** Macro-level **GPSM using NUTS 3 or other spatial data**	Macro-level **Naïve Baseline Comparison**
Rigour	**Approach A is rigorous** but controlling selection bias will require an extra effort.	**Approach B is not rigorous.** It only gives a general idea about trends in changes and no exact findings.
Reliability	**Approach A is partly reliable.** Its degree of reliability depends on sample size and sample sizes are not expected to be sufficiently large to safeguard the exactness of findings.	**Approach B is partly reliable**. It can show the trends of change well but, it produces a rather fuzzy estimate of the net direct effect.
Robustness	**Approach A is partly robust.** Its robustness depends on sufficient sample size.	**Approach B is not robust.** It can only be used to forecast the trends in expected changes.
Validity	**Approach A is partly valid.** Its validity depends on the structure of the sample. In general, sample selectivity is expected to be very low at least in the livestock/manure management sector/activity.	**Approach B is partly valid.** The method can depict the trends in changes but not an exact estimate.
Transparency	**Approach A is transparent**, but special attention must be paid to choosing 'good' control variables, sound instruments if the Instrumental Variables (IV) approach is chosen or a good matching algorithm in constructing the counterfactual.	**Approach B is transparent,** especially when all possible sources of information and secondary data are available and collected.
Credibility	**Approach A is credible.** Sample selection is restricted and is not expected to have serious impacts.	**Approach B is credible** for producing an estimate of the 'broad quantitative range' of effects.

Practicability	**Approach A is practical** for small sample sizes, intense pre-processing of monitoring data, or if combined with other similar studies.	**Approach B is practical**, since it can provide a quick assessment of gross RDP's effects.
Cost-effectiveness	**Approach is cost-effective** for small sizes and intense pre-processing of monitoring data	**Approach B is very cost effective** on the expense of reliability and rigour.

4.2.5 Dos and don'ts

Dos

- Contact the focal point (any other relevant stakeholder) for reporting national GHG and ammonia emissions of your country and ask for the latest available figures for the I.07 and/or additional indicators, even if they are unpublished.

- Ask whether a regional unpublished dataset exists and if the methodology (equations) can be mechanically applied to regional data in case of regional RDP.

- Build database, but before starting to explore whether you can develop synergies with other evaluations that may have the same approach, e.g., evaluation of water quantity, water quality and especially GNB, soil quality and soil erosion.

- Review your IACS/LPIS database, existing farm holding sampling frames specifically addressing emissions (at least manure handling and fertilisation), existing GIS maps, locate data gaps and get a first-hand idea of the blend of measures used to reduce agricultural emissions.

- Clarify the criteria (eligibility and location) that would categorise a farm holding to the control group.

Don'ts

- Build a regional database for I.07 which is based on another methodology (tier) than that used by the national inventory report.

29

4.3 CAP common impact indicators I.08 and I.09

The following sections of the Technical Annex are related to Chapters 2.4 and 2.5 in PART II of the Guidelines.

4.3.1 Additional indicators (examples)

Additional indicators can also be used in the assessment of RDP effects on biodiversity and HNV farmland in case the common CAP impact indicators cannot provide sufficient information. Examples are provided in the Table 6 below.

Table 6.   Examples of additional indicators related to biodiversity and HNV farmland

Indicator	Unit of measurement	Use of additional indicator	Data sources and frequency of collection
Number of flora and fauna species on contracted land[26].	- Number of flora and fauna species. - % of increase/ decrease of flora/fauna on contracted land and among them those endangered.	Additional result indicator to address gaps between common result and impact indicators and to allow a broader biodiversity assessment.	- Biodiversity monitoring data on participating and non-participating land in the timeframe of the programme period under consideration. - National habitat biodiversity monitoring programmes.
Number of farmland bird individuals[27].	Number of farmland bird individuals.	- To address potential data gaps of the FBI, to assess micro-level impacts. - More robust counterfactual assessment at micro level compared to FBI, as the unit of analysis is linked to distinct parcels of contracted or not contracted areas.	- Regularly collected bird census data on participating and non-participating land in the timeframe of the programme period under consideration. - National habitat biodiversity monitoring programmes.
Singing males of corncrakes (example of individual bird species indicator)[28].	Number of singing corncrake males.	- Example for additional result indicators to assess specific biodiversity impacts in addition to FBI – no application on its own. - Individual species trends help to understand mechanisms that drive trend in farmland birds.	- Regularly collected data on singing males of corncrakes, land cover data and agricultural land-use data for participating and non-participating land. - National habitat biodiversity monitoring programmes.
Bumblebee indicator[29].	- Number of bumblebees. - Number of bumblebee species.	Supplements the farmland bird indicator with an additional taxonomy group with different habitat niches and behaviours and also provides evidence on changes in the quality of HNV farmland.	- Long term bumblebees monitoring data on participating and non-participating land / farms. - Specific national habitat and biodiversity monitoring programmes.
Population trends of agriculture related butterfly species.	Rate of change in the relative abundance of agriculture related butterfly species.	Supplements the farmland bird indicator with an additional taxonomy group with different habitat niches and behaviours and also provides evidence on changes in the quality of HNV farmland.	Long term monitoring data on population trends of agriculture specific butterfly species on participating and non-participating land/farms. Specific national habitat and biodiversity monitoring programmes.

---

[26] More information can be found at: European Evaluation Helpdesk for Rural Development (2016) Assessments of RDP Results – how to prepare reporting in 2017. Annex 11 - Fiches for answering Common Evaluation Questions for rural development programmes 2014-2020 (CEQ 1 – 21).

[27] More information can be found at: 'Methodological Handbook for the evaluation of environmental impacts of RDPs',. ENVIEVAL- Project: https://www.envieval.eu/dissemination-plattform/deliverables/, Morkvenas et al., (2015).

[28] More information can be found at: 'Methodological Handbook for the evaluation of environmental impacts of RDPs',. ENVIEVAL- Project: https://www.envieval.eu/dissemination-plattform/deliverables/, Morkvenas et al. (2015).

[29] More information can be found at: European Evaluation Helpdesk for Rural Development (2016) Estonian Case Study – the Evaluation of Agri-environment Schemes' Biodiversity Objectives. Rural Evaluation Fact Sheet.

Different data availabilities of various additional biodiversity indicators across the EU Member States imply that the selection of additional indicators needs to be reviewed on a case-by-case basis. At EU level data of the FBI are available. Data for the number of farmland bird individuals is the baseline data for the FBI. Data collection (monitoring) standards of the common bird species are set by the European Bird Census Council. In most cases, however, the FBI does not have such a good national and regional coverage, or the data do not coincide with areas under specific measures. Therefore, the index should be integrated with other previous or ongoing bird monitoring data, if they exist, or combined with data of existing common bird monitoring and special measure-specific studies.

Data on rather specific biodiversity indicators, such as the corncrake singing males, are not gathered systemically, but are available in countries which report on the conservation status of corncrakes, according to the reporting requirements for the EU Birds Directive implementation. Biodiversity data collected for the reporting of the EU Birds and Habitat Directive are suitable for RDP assessments if adequate coverage of participants and non-participants is given, and data are available from before and after implementation of RDP measures. Other necessary data are land-cover data (e.g. CORINE land cover), agricultural land-use data (IACS/LPIS) and farm surveys. The resolution of Land Use and Cover Area frame Survey (LUCAS), land cover data can be insufficient for micro level evaluations. EU databases are useful at EU level but have several limitations which constrain their application for national/ regional HNV impact evaluations.[30] National and specific, regional and local land cover maps and biodiversity monitoring programmes from different organisations (covering different aspects of biodiversity) play a key role in providing the data necessary for evaluating biodiversity impacts of RDPs. In addition, other alternatives for CORINE, for land cover data, such as freely-available spatial data (e.g. Google Earth) and remote-sensing data (e.g. Copernicus Programme) can be considered.

4.3.2 Approach A and B for I.08 – Farmland Bird Index

Micro level assessment

Access to data and its quality & creation of a consistent database and data infrastructure

In this step, assessments of **quantity and quality of data** are carried out to decide upon evaluation approach to use. The task also involves checking if data characteristics are appropriate for the implementation of the **chosen evaluation approach (Approach A/B)** and available to assess biodiversity impacts of the RDPs. Some key questions to consider:

- Is the amount and characteristics of data appropriate to implement a robust evaluation approach for biodiversity impact evaluation?
- What options are available to construct a counterfactual for evaluating biodiversity impacts?
- Does the uptake of the evaluated measure(s) and availability of spatially explicit biodiversity monitoring data allow constructing a (or several) control group(s)?
- To what extent do I have data on other factors influencing farmland biodiversity?
- Do I have data for the selected biodiversity indicator for different points in time (before and after) for participants and non-participants?
- Do I need to collect new primary data through statistical sampling?
- Is there a need for specific processing tasks to improve the quality of the farm survey/biodiversity monitoring data?
- What are the implications for the costs of the evaluation and its potential performance?

The assessment of data quality and quantity includes **tests on: (i) the scope for increasing the quantity of data** (e.g. number of observations) **to assure a better representativeness** of the results

---

[30] For more information see: Evaluation Helpdesk WD: Practices to Identify, Monitor and Assess HNV Farming in RDPs 2014-2020, (2016).

and creation of robust comparison groups of participants and non-participants, and **(ii) whether data pre-processing may be required**.

Firstly, statistical analysis, statistical sampling techniques and expert analysis are available **to review the representativeness of the data**. In particular for data on biodiversity additional indicators it is advised that the evaluator will consult local data providers of biodiversity monitoring programmes to verify the representativeness of their data. Background, purpose and sampling strategy of the existing monitoring data needs to be assessed to decide their suitability, as in most cases biodiversity monitoring programmes were not designed for RDP evaluation and thus, might not provide representative coverage of different types of farmland as well as types of participant and non-participants. Details of geographical site location, land use, farm management, habitats, landscape features etc., within and around the survey spots need to be available for use as co-variates in the spatial data analysis, to account for variation not directly explained by the measure.

Secondly, if **data gaps** which **restrict the application of advanced statistics-based evaluation approaches (Approach A)** are identified**,** opportunities for collecting additional data or/ and applying additional indicators with better data availability need to be explored. Additional data collection can include specific farm surveys and specific biodiversity monitoring sampling. Samples have to be representative in terms of habitat and geographical location. Therefore, the best way to gather representative samples is to use a large number of random plots. However, while the use of volunteers to conduct the biodiversity monitoring has been successfully applied, options for additional sampling will often be restricted by the time and budget available in evaluations. Alternatively, a set of specific geographical and thematic biodiversity case studies can be conducted to address data gaps.

Thirdly, the **formation of comparison groups** is particularly important when self-selection of programme participation is likely. When farmers are not randomly assigned as participants to the evaluated programme (e.g. through specific spatial or biophysical settings targeted by biodiversity measures), a simple comparison of programme participants and non-participants will lead to biased impact estimation of an unknown magnitude and direction.

RDP's monitoring databases and especially IACS/LPIS are a useful starting point for drawing samples for parcels on participating and non-participating farms. However, the construction of comparison groups is data intensive and requires substantial biodiversity monitoring data with survey points at a suitable spatial distribution for participants and non-participants. Spatial aspects of the indicator species and the use of existing monitoring programmes are key factors determining the feasibility of constructing comparison groups in the counterfactual assessment of relevant measures under the focus area 4A Biodiversity. Temporal or spatial scarcity of data of non-participants can particularly hinder the construction of comparison groups and needs to be carefully reviewed in an assessment of existing biodiversity monitoring programmes.

Lastly, a **range of different observable characteristics,** which can explain the participation of farms (and parcels), need to be considered in the design of comparison groups including eligibility of participation, participation in previous programming periods different farms, spatial and bio-physical characteristics (examples can be the degree of naturalness or proximity to protect areas, groundwater levels, slope and size). For some of these characteristics information can be drawn from IACS / LIPS and FADN. As a consequence, multiple comparison groups might be needed depending on data availability and selected econometric method.

Selection of a counterfactual option and micro-level method

As the Approach A at micro level, assuming sufficient data availability and quality, joint **PSM/DiD**[31] can be applied assessing net-effects on the FBI and selected additional biodiversity indicators. Steps for application of PSM/DiD are described in PART II, Chapter 2.1 in the box called 'How to construct the control groups at the micro level? (2nd layer)'– Quick guide #2.

The integration of a specific environmental method (e.g. particular environmental simulation models) in the evaluation approach is not necessary, as the indicators and set of variables used for the statistics-based technique are sufficiently self-explanatory.

But as stated above, data availability, sampling and resolution of biodiversity monitoring programmes can constrain the application of such data intensive approaches with the FBI (and other biodiversity indicators). In cases of insufficient farm structural, spatial and bio-physical data for variables that could explain the participation advanced econometric methods, such as Propensity Score Matching, can't be applied cost-effectively, and **ad-hoc approaches to sample selection**[32] have to be used instead. In such cases, less robust ad-hoc pairwise comparisons, or multiple comparison groups form the basis of a statistical analysis of trends on treatment parcels and control parcels using the DiD approach. Such an approach is considered to be an example of an **alternative evaluation approach in case of data gaps**.

Good practice of evaluating biodiversity impacts of RDPs at micro level also entails additional analysis with the FBI – or similar indicators using FBI data - on a smaller scale. Population trends of separate species, or particular prominent and relevant bird species, can be analysed to identify whether rapid changes in populations during the programming period can be identified. Individual species' trends help to understand mechanisms that drive trends in farmland birds. Good practice shall also include the use of biodiversity indicators reflecting changes in important taxonomic groups other than birds, complementing the use of the FBI. The use of case studies is advised in cases of a) large thematic or geographical differences in data availability, b) in-depth study of a wider range of biodiversity impacts and/ or c) in-depth studies of particular measures or sub-measures of specific interest.

Net impact assessment at micro level

For **Approach A** the application of PSM/ DiD enables the measurement of deadweight loss effects by calculating the average treatment of the treated.

As for **Approach B,** the calculation of the average treatment of the treated is less robust with an ad-hoc approach in the construction of comparison groups. On the other hand, the careful design of the groups through a wise choice of control variables in combination with less data intensive statistical/ econometric analysis can reduce the selection bias to some extent. However, deadweight and displacement[33] effects may be difficult to quantify and may, at best, be addressed in a qualitative and contextual manner. Steps for application of PSM/DiD are described in PART II, Chapter 2.1 in the box called 'How to assess RDP net effects? (3rd layer).' – Quick guide #4.

---

[31] For more information and methodological details see: 'Counterfactual impact evaluation of EU rural development programmes - Propensity Score Matching methodology applied to selected EU Member States.', Volume 1 micro level approach. EUR 25421 EN, Michalek, J., (2012).

[32] For more information and methodological details see: 'Methodological Handbook for the evaluation of environmental impacts of RDPs', ENVIEVAL-Project: https://www.envieval.eu/dissemination-plattform/deliverables/, Morkvenas et al. (2015).

[33] Guidelines for ex post evaluation of 2007-2013 RDPs, Chapter 4.2.3.

Macro level assessment

Access to data and its quality & creation of a consistent database and data infrastructure

Similarly to the micro-level evaluation, assessments of the quantity and quality of data need to be carried out to check if their characteristics are appropriate for the implementation of the evaluation approach to assess biodiversity impacts of the RDPs at macro level. Some key questions to consider:

- Is the amount and characteristics of data appropriate to implement a robust evaluation approach for biodiversity impact evaluation at macro level?
- Do I need to harmonise the spatial resolution of different data sets?
- Is there a need for specific processing tasks to improve the quality of the farm survey/biodiversity monitoring data?
- What additional sources of information do I need to extrapolate from micro-level results to macro level?
- What are the implications for the costs of the evaluation and its potential performance?

For **Approach A** relatively large samples are needed to achieve representativeness at regional level (e.g. NUTS 3 level). In many cases however, the FBI will not have a good territorial coverage or the data will not coincide with areas under specific measures. In such cases, for **Approach B,** the use of other previous or ongoing national, regional and local bird monitoring data, should be investigated as well as ad hoc and highly replicated field studies, including pair-wise comparisons with control sites.

If spatial data used at macro level are at different spatial resolutions (e.g. spatial support for economic actors at a municipal scale, biodiversity data at parcel or landscape scale, soil data for individual soil units) an extra step is required prior to analysis. At this test the spatial resolution needs to be harmonised either through up-scaling or down-scaling methods to a single resolution.

Selection of counterfactual option, macro-level method and net assessment

Similarly to micro level evaluations, and assuming sufficient data availability and quality, for approach A the joint PSM/DiD[34] can be applied assessing net-impacts on the FBI and selected additional biodiversity indicators. Funding intensities of bio-geographical areas or at regional level could for example be used for matching processes to define comparison groups.

In addition, spatial econometric models, provide a suitable method to assess biodiversity impacts of RDPs in bio-geographical areas (different agricultural habitats) or at regional level, which allows for the incorporation of counterfactuals through analysing areas or regions with different spending on the measures and different development trajectories of biodiversity.[35] Spatial econometrics is specifically able to disentangle the external impacts of other intervening factors and can be applied to different spatial levels (e.g. NUTS 1, but more suitable NUTS 3). However, spatial econometric models require a comprehensive database of land use, farm management and characteristics, and biodiversity data at (farm and) regional level. Data processing requirements are substantial which demand specific methodological skills and interest from the evaluator.

As stated above, data availability and sampling, and resolution of biodiversity monitoring programmes can constrain the application of such advanced and data intensive modelling approaches with the FBI (and other biodiversity indicators). In such cases bottom-up approaches upscaling micro level findings, based on a sufficient number of case studies can be used to extrapolate to the macro scale, and are considered as **an alternative approach in case of data gaps.**

---

[34] For more information and methodological details see: 'Counterfactual impact evaluation of EU rural development programmes. Propensity Score Matching methodology applied to selected EU Member States', Volume 2 a regional approach. EUR 25419 EN., Michalek, J. (2012).

[35] For more information and methodological details see: 'Using spatial econometrics in impact assessment', SPARD deliverable D4.5. The Hague, LEI-Wageningen UR, Reinhard A. J., Linderhof V. (2013).

Given the complexity of developing robust comparison groups for evaluating biodiversity impacts, deadweight and displacement effects may be difficult to quantify and may, at best, be addressed in a qualitative and contextual manner, or by demanding multivariate approaches.

Micro-macro consistency and validation

For the net impact assessment, it is important that the results of both micro and macro-level assessments are consistent, i.e. the results of these assessments show the same trend in relation to impact, even though the evaluator used different indicators or even different methods for the assessments. More information on micro-macro consistency check can be found in PART II, Chapter 2.1.

Pros and cons of using the Approach A

Important feature	Pros	Cons
Propensity Score Matching is a robust method for dealing with selection bias and – used in combination with a DiD approach - to quantify net-effects. It however requires large sample sizes of participants and non-participants for which biodiversity data are available to ensure that the largest proportion of RDP participants can be matched to a non-participant.	Robust method for addressing selection bias.	Biodiversity data required for large samples of participants and in particular non-participants will in many cases not be available in 2018/2019 and thus exclude an application. Intermediate econometric skills are required.
Spatial econometrics can assess biodiversity impacts of RDPs in bio-geographical areas (different agricultural habitats) or regional level Counterfactuals can be incorporated through analysing areas or regions with different spending on measures and different development trajectories of biodiversity.	Robust method which is able to disentangle the external impacts of other intervening factors and can be applied to different spatial levels.	- High data requirements. - Specific methodological and advanced econometric skills are required.

Pros and cons of using Approach B

Important feature	Pros	Cons
Statistical analysis of ad-hoc pairwise comparisons, or multiple comparison groups using DiD can be differentiated by known factors and observables, to guarantee reduced biases in the evaluation results without the high data requirements of PSM.	-Relatively simple application. - Less data demanding. - Transparency is given, as main steps and elements can be clearly explained and followed.	Will not produce exact findings, but rather indicate the direction of change, as other intervening factors are not fully considered.

Preconditions for applying Approach A and B

> The crucial decision is whether there will be sufficient biodiversity monitoring data to use advanced statistics based evaluation approaches, such as PSM and DiD to quantify net biodiversity impacts. In this respect the following shall be clarified before contracting the evaluator:
>
> - Can you robustly use statistics-based methods to quantify biodiversity net-effects of the evaluated measure(s), or do you need to consider alternative (ad-hoc) options to consider sample selection issues?
> - What options are available to construct a counterfactual for evaluating impacts on biodiversity and HNV farming?
> - Does the uptake of the evaluated measure(s) and availability of spatially explicit biodiversity monitoring data allow constructing a (or several) control group(s)?
> - To what extent do I have data on other factors influencing farmland biodiversity?
> - Do you have data for the FBI, HNV farming and selected additional biodiversity indicator for different points in time (before and after) for participants and non-participants?
> - Do you need to collect new primary data through a farm survey?
> - What are the implications for the costs of the evaluation and its potential performance?
> - Will you require specific skills and proven experience in applying econometric and qualitative methods for biodiversity assessments from those who will develop and apply the models and methods?

4.3.3 Approach A and B for I.09 HNV farming

Micro level assessment

Access to data and its quality & creation of a consistent database and data infrastructure

Similarly as for the biodiversity impact indicator in case of HNV farming the **assessments of the quantity and quality of data** shall also take place **before deciding on the evaluation approach**. The same key questions, as for the impact indicator I.08 need to be considered.

The HNV farming indicator requires georeferenced data on land cover and land use with sufficient details to guarantee:

- the assessment of semi-natural features;
- the level of farming intensity and presence of wildlife species;
- the possibility of comparing participants and non-participants (e.g. IACS/LPIS database).

Assessing existing data includes tests on: (i) the scope for increasing the quantity of data (e.g. number of observations) to ensure a better representativeness of the results and creation of robust comparison groups of participants and non-participants, and (ii) whether data pre-processing may be required.

Firstly, for statistical analysis, statistical sampling techniques and expert analysis are available for reviewing the representativeness of the data. EU, national and regional databases were not designed for monitoring tendencies in HNV farming and need to be complemented with more detailed sample surveys. If specific HNV monitoring programmes exist in the Member State in question, it is advised that the evaluator will consult local data providers to verify the representativeness of their data. Background, purpose and sampling strategy of the existing monitoring data needs to be assessed to decide their suitability for the use of creating robust comparison groups, as in most cases HNV monitoring programmes were not designed for RDP evaluation and might thus not provide

representative coverage of different types of participants and non-participants. Details of geographical site location, land use, farm management, habitats, landscape features etc. within and around the survey spots need to be available for use as co-variates in the spatial data analysis, to account for variation not directly explained by the measure.

Secondly, if data gaps which restrict the formation of comparison groups and the application of advanced statistics-based evaluation approaches are identified, opportunities for collecting additional data and/or applying proxy indicators with better data availability need to be explored. Additional data collection can include specific farm surveys and specific biodiversity monitoring sampling. Samples have to be representative in terms of farming systems, habitat and geographical location. Options for additional sampling will often be restricted by time and budget available in evaluations.

RDP's monitoring databases, especially IACS/LPIS, are a useful starting point for drawing samples for parcels on participating and non-participating farms. However, the construction of comparison groups is data intensive and requires substantial HNV monitoring data with survey points covering the different indicators for semi-natural vegetation, mosaic of low-intensity agriculture and biodiversity aspects (wild species and habitat of conservation concern), at a suitable spatial distribution for participants and non-participants. Spatial aspects of the indicators and the use of existing monitoring programmes are key factors determining the feasibility of constructing comparison groups in the counterfactual assessment of relevant measures, under the focus area 4A Biodiversity. Temporal or spatial scarcity of data of non-participants can hinder the construction of comparison groups and needs to be carefully reviewed in an assessment of existing biodiversity monitoring programmes.

Another major data limitation concerns the lack of information on the extent of semi-natural features on the farms, and more generally in terms of land cover.[36] Semi-natural vegetation plays a major role in the provision of green infrastructures that significantly the biodiversity values of a farmland area. Up until now, other indicators have not proven to be sufficiently good proxy indicators. On the other hand, the increasing availability of data concerning large and small patches of perennial vegetation detected in fine-resolution satellite images, should increase the reliability of land cover in agro-ecosystems at reasonable monitoring costs.[37]

Lastly, as explained in Annex 4.3.2, the formation of comparison groups is particularly important when self-selection of programme participation is likely. A range of different observable characteristics, which can explain the participation of farms (and parcels), need to be considered in the design of comparison groups. The characteristics include eligibility of participation, participation in previous programming periods, and different farm, spatial and bio-physical characteristics (including land cover, parcel size, proximity between participant and non-participant, etc.). As a consequence, multiple comparison groups might be needed, depending on data availability and selected econometric method.

Selection of a counterfactual option and micro-level method

As **evaluation approach** at micro level, an advanced technique - joint PSM/DiD[38] can be applied assessing net-effects on HNV farming and selected additional biodiversity indicators. This technique is applied, if the sample of farms has a reasonable representativeness of participants and non-participants, for example through combining IACS/LPIS data with HNV monitoring data. The application

---

[36] For more information and methodological details see: 'Do agri-environmental schemes contribute to High Nature Value farmland? A case study in Emilia-Romagna (Italy)', Ecological Indicators, 59, 62 – 69, Viaggi D., Signorotti C., Marconi V., Raggi M., (2015).

[37] For more information and methodological details see: 'Semi-natural vegetation in agricultural land: European map and links to ecosystem service supply.', Agronomy for Sustainable Development 35(1): 273-283, García-Feced C., Weissteiner C., Baraldi A., Paracchini M., Maes J., Zulian G., Kempen M., Elbersen B., Pérez-Soba M., (2015).

[38] For more information and methodological details see: 'Counterfactual impact evaluation of EU rural development programmes - Propensity Score Matching methodology applied to selected EU Member States.', Volume 1 micro level approach. EUR 25421 EN., Michalek J., (2012).

of PSM/DiD enables the measurement of deadweight loss effects by calculating the average treatment of the treated.

In cases of limited data availability for the evaluation in 2019, alternatively, less robust ad-hoc pairwise comparisons or multiple comparison groups[39] can form the basis of statistical analysis of trends on treatment parcels and control parcels using the DiD approach. The calculation of the average treatment of the treated is less robust with an ad-hoc approach in the construction of comparison groups, but careful design of the groups through a wise choice of control variables in combination with less data intensive statistical/econometric analysis can reduce the selection bias to some extent. However, deadweight and displacement effects may be difficult to quantify and may, at best, be addressed in a qualitative and contextual manner.

In cases where the sample of participating and non-participating farms in the HNV survey is too small to conduct statistical analysis across comparison groups, case study assessments of specific areas or small groups of farms can be conducted as a last resort.

Net impact assessment at micro level

Assessing net-impacts on HNV farming involves, however, the use of a combination of different indicators, covering the different aspects of HNV, to gather an understanding of how HNV farming is evolving and the impact of RDPs upon it. Expert judgement is then used to assess the overall impact and role of RDPs.

Macro level assessment

Access to data and its quality & creation of a consistent database and data infrastructure

Similarly to the micro level assessment, assessments of quantity and quality of data need to be carried out to check if their characteristics are appropriate for the implementation of the evaluation approach to assess the impacts of RDPs on HNV farming at macro level. The key questions outlined for the data assessment for the FBI also apply here.

At macro level, for using the **Approach A**, the spatial analysis concerning participants and non-participants is applicable if the IACS/LPIS databases are available at cadastral level, and a consistent geodatabase is also needed. Two sources of data from IACS are particularly important: a LPIS GIS database with aerial photo interpretation and the Farm sheet registry with an extensive range of information on agriculture holdings (e.g. personal data, farm location, n. of applications for admission to EU funding and national aids, measure implemented under pillar 1 and 2 of the CAP, etc.). In addition, bio-physical data and available GIS maps on various items concerning land cover, nature conservation and biodiversity can be added. As for GIS systems, they can be created based on cadastral map sheet as territorial unit of analysis.

Selection of a counterfactual option, macro-level method and net assessment

For Approach A, the spatial econometric models provide a suitable method to assess the impacts of RDPs on HNV farming at macro level. This allows for the incorporation of counterfactuals through analysing regions with different spending on measures and different development trajectories of HNV farming.[40] Spatial econometrics are specifically designed to disentangle the external impacts of other intervening factors and can be applied to different spatial levels (e.g. NUTS 1, but more suitable NUTS 3). However, spatial econometric models require a comprehensive database of land use, farm management and characteristics, and biodiversity data at farm and regional level. Data processing

---

[39] For more information and methodological details see: 'Methodological Handbook for the evaluation of environmental impacts of RDPs.', ENVIEVAL-Project: https://www.envieval.eu/dissemination-plattform/deliverables/, Morkvenas et al. (2015).

[40] For more information and methodological details see: 'Using spatial econometrics in impact assessment.', SPARD deliverable D4.5. The Hague, LEI-Wageningen UR., Reinhard A. J., Linderhof V., (2013).

requirements are substantial, demanding specific methodological skills and interests from the evaluator/user.

As stated above, if gaps in data availability and sampling do not allow the application of spatial econometric models, **bottom-up approaches based up-scaling micro level results** can be used as an **alternative approach** to extrapolate onto the macro scale. These approaches will be linked with GIS analysis to enable explicit spatial data analysis. The process of upscaling to macro level will be carried out with the use of statistical technique, such as kriging[41]. This technique, that produces an estimation of the underlying surface by a weighted average of the data, where weights decline with distance between the point at which the surface is being estimated, and the locations of the data points. The HNV identification approach is based on the characterisation of the functional territorial unit (cadastral sheets) in relation to the four specific indicators.

The **alternative macro level evaluation approach** to assessing the net impact, is based on a **naïve counterfactual approach**. More precisely, the identification of the HNV area at the macro level is done with the use of a geostatistical method, which provides the formulation of separate indicators, reflecting the different types of HNV for example:

- semi-natural vegetation through the identification of semi-natural vegetation land cover classes from Land Cover-LPIS datasets and the calculation of their area with GIS software;
- ecological diversity through the Shannon Index.

Through standardisation, weighing and aggregation of the indicators, it is possible to compute the composite indicator. This indicator allows mapping the overall nature value ordering of cadastral map sheets according to the composite indicator value. Therefore, the identification of HNV is based on the 25% (first quartile) of cadastral map sheets with the highest composite indicator value as HNV at the regional scale. Finally, the area treated by RDP measures can be measured from the overlay of the areas taken from the RDP monitoring system and the percentage of area located inside HNV.

Displacement effects may be difficult to quantify and may, at best, be addressed in a qualitative and contextual manner.

Micro-macro consistency and validation

For the net impact assessment, it is important that the results of both micro and macro level assessments are consistent, i.e. results of these assessments show the same trend in relation to impact, even though the evaluator used different indicators, or even different methods, for the assessments. More information on micro-macro consistency check can be found in PART II, Chapter 2.1.

Steps in the application of Approach B, pros and cons in using this approach (can also be found in Annex 4.3.2).

4.3.4 Adequateness of suggested evaluation approaches

The above described evaluation approaches are discussed below as regards their adequateness in fulfilling the evaluation quality criteria: rigour, reliability, robustness, validity, transparency, credibility, practicability and cost effectiveness. The definitions of the quality criteria are provided in Table 2.

---

[41] Kriging is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z-values. Retrieved from ArcMap http://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.htm#ESRI_SECTION1_E112B7FAED26453D8DA4B9AEC3E4E9BF

Table 7.   **Adequateness of the proposed evaluation approaches for the assessment of CAP common impact indicator I.08 – Farmland Birds Index (FBI) and I.09 High Nature Value (HNV) farming**

Quality criteria	Approach A (optimal) Micro-level **Propensity Score Matching (PSM) and Difference in Differences (DiD)** Macro-level **PSM and DiD and special econometric models at bio-geographical areas**	Approach B (alternative) Micro-level **Statistical analysis of ad-hoc pairwise comparison groups using DiD** Macro-level **Up-scaling micro-level findings accompanied by qualitative assessment**
Rigour	**Approach A is rigorous,** since it delivers exact findings on net-impacts.	**Approach B is partly rigorous.** Ad-hoc approaches to sample selection will only indicate the trends of change, as other intervening factors are not fully considered with the approach.
Reliability	**Approach B is highly reliable.** If all analytical steps are repeated, similar results will be generated in identical conditions.	**Approach B is partly reliable,** since other intervening factors are not considered in the assessment which is affecting the assessment findings and reducing reliability.
Robustness	**Approach A is robust,** but depends on sample sizes and assuming there is sufficient availability of robust biodiversity datasets for participants and non-participants.	**Approach B is partly robust**, it depends on the design of the comparison groups and the extent to which sample selection issues and known factors, and observables are considered.
Validity	**Approach A is valid.** It delivers logical and accurate evaluation findings, and enables sound conclusions to inform policy recommendations.	**Approach B is partly valid.** The accuracy of the approach is restricted, as the calculated effects will not fully reflect isolated programme effects.
Transparency	**Approach A is partly transparent.** Transparency can be reduced for stakeholders due to complex nature of the approach.	**Approach B is transparent.** The main steps and elements can be clearly explained and followed.
Credibility	**Approach A is highly credible**, as causality is well founded and programme effects are isolated.	**Approach B is partly credible.** Results are based on robust theoretical causality but calculated effects might not fully reflect isolated programme effects.
Practicability	**Approach A is partly practical.** Insufficient availability of biodiversity monitoring data, for participants and non-participants, can restrict the application of the approach.	**Approach B is partly practical,** but it can provide an acceptable way of assessing biodiversity impacts under limited data availability.
Cost-effectiveness	**Approach is cost-effective.** Although the approach is costlier, it delivers rigorous and robust results of net-impacts, if the required data are available. Under such circumstances the approach is cost-effective.	**Approach B is cost effective.** It can provide a cost-effective way of assessing biodiversity impacts under limited data availability.

4.3.5 Dos and don'ts

Dos

- Consider selecting additional indicators providing additional evidence on biodiversity impacts of the RDP.
- Carefully review available data on monitoring biodiversity and contact relevant monitoring organisations.
- Differentiate *ad-hoc* approaches through careful design of pairwise comparisons and multiple comparison groups, by known factors and observables. These groups could be spatial neighbourhood, proximity to protected areas, or sub-groups of beneficiaries. This way, using the DiD approach, biases should be reduced in the evaluation results.
- Select an evaluation approach which is consistent with the quantity and quality of available data for beneficiaries and non-beneficiaries.

Don'ts

- Rely only on data of the Farmland Bird Index in case of insufficient sampling points in the RDP region.
- Apply simple average aggregated comparisons of beneficiaries and non-beneficiaries, if using ad-hoc group comparisons.

4.4   CAP common impact indicators I.10 and I.11

The following sections of the Technical annex are related to Chapter 2.6 in PART II of the Guidelines.

4.4.1 Additional indicators (examples)

The indicator I.10 refers to the volume of water which is applied to soils for irrigation purposes only, irrespective of the water source, surface or groundwater. We suggest that the evaluator be supported and gets a better insight by examining certain additional indicators, such as the total water consumed in agriculture and the water exploitation index, or specific indicators serving unique RDP interventions. The latter refers, for example, to RDPs that do not target on farm water reductions, but target water savings through infrastructure in the storage and distribution network.

Furthermore, we propose additional indicators to capture the risk of water pollution by pesticides and phosphorous since these can, potentially, be a long-term issue. Additional indicators are summarised in the Table 8 below:

Table 8.   Examples of additional indicators related to water abstraction and water quality

Indicators	Measurement unit	Use of additional indicator	Data sources and frequency of collection
-		- Additional indicators linked to I.10 - Water Abstraction	
Water abstraction in agriculture (total)	'000 of m^3 abstracted. per year and per River Basin District for the WFD and per country for OECD/Eurostat.	Water abstraction in agriculture measures all the water quantities abstracted for agricultural use including the quantities lost in the storage and distribution network.	Water abstraction data are collected for monitoring the implementation of the Water Framework Directive (WFD).[42] Yearly data on freshwater resources, water abstraction and use, are also collected by Eurostat by means of the OECD/Eurostat Joint Questionnaire - Inland Waters.[43] WFD Data do not cover administrative regions.
The Water Exploitation Index (WEI) and the Regional Water Exploitation Projection	%	This is a Sustainable Development Indicator (SDI) and presents the annual total fresh water abstraction differentiated by surface and groundwater, as a percentage of its long-term average available water (LTAA) from renewable fresh water resources.	Eurostat[44] reporting is voluntary, and many Member States do not report (e.g. Belgium, Ireland, Italy) or report very sparingly (e.g. Austria, Germany, Finland). Data do not cover regions. The Regional Water Exploitation (RWE) as a fraction of projected available water in 2030, is available by JRC at its water portal.[45] WEI calculated yearly, and RWE has been calculated only once.
Efficiency of the water logistics network	'000 m3 of water savings.	Certain RDPs have chosen to implement actions to save water that is lost in the storage and distribution network.	RDP own sources and monitoring data.
Sustainably irrigable areas	ha of areas made irrigable.	Certain RDPs have chosen to implement infrastructure that will deliver irrigation water under conditions that will allow the adoption of sustainable on farm irrigation methods.	RDP own sources and monitoring data.
-		- Additional indicators linked to I.11 - Water Quality	
Mineral fertiliser consumption	million tonnes of nutrients.	This is AEI 5 (Eurostat indicator). The consumption of the nutrients nitrogen (N) and phosphorous (P) as mineral fertilisers by agriculture is an indicator of nutrient availability.	Estimated consumption based on the sales of mineral fertiliser in the EU-28 from Fertilizers Europe.[46] Data

---

[42] Waterbase-Water Quantity can be accessed at: https://www.eea.europa.eu/data-and-maps/data/waterbase-water-quantity-9 and the WFD implementation authorities of each Member State can be found at: http://ec.europa.eu/environment/water/water-framework/links/index_en.htm http://rod.eionet.europa.eu/contacts?roleId=eionet-nfp-mc For the reporting obligation of WISE SoE - Water Quantity (WISE-3) evaluators may ask information from their national focal point contacts at (http://rod.eionet.europa.eu/contacts?roleId=eionet-nfp-mc) and their national reporting contacts at: http://rod.eionet.europa.eu/contacts?roleId=reportnet-awp-wise3-reporter

[43] Data and metadata accessed at: http://ec.europa.eu/eurostat/en/web/products-datasets/-/ENV_WAT_ABS

[44] Eurostat data can be accessed at: http://ec.europa.eu/eurostat/web/products-datasets/-/sdg_06_60

[45] JRC water portal accessible at: http://water.jrc.ec.europa.eu/waterportal

[46] Fertilizers Europe website at: http://www.fertilizerseurope.com/

Indicators	Measurement unit	Use of additional indicator	Data sources and frequency of collection
			are at country level excluding Croatia. Data are collected yearly with a time lag of around 2 years.
Pesticide pollution of water	%	This is AEI 27.2. Measuring groundwater with pesticide concentrations above Environmental Quality Standards (EQS) and rivers with annual average pesticide concentrations above EQS.	Water quality data are collected for monitoring the implementation of the Water Framework Directive (WFD) or Nitrate Directive. Yearly data have a time lag of around two and a half years and do not cover administrative regions.[47]
Risk of pollution by phosphorus	Categorical measurement in five P retention classes from very weak to very strong.	This is part of AEI 16. The measure of phosphorous retention capacity, as documented by the indicator's fiche in 2005, is important for planning agri-environment programmes. Sorption of P is important because it captures the ability of soils to change soluble phosphate to less soluble forms, by reacting with inorganic or organic compound of the soil so that P becomes immobilised.	P retention has been calculated by Bomans et al (2005) only once.[48]

---

[47] Waterbase - Water Quality can be accessed at: https://www.eea.europa.eu/data-and-maps/data/waterbase-water-quality
[48] Bomans et al. (2005). Methodology and data are available in the report which is accessible through the EU's website at: http://ec.europa.eu/environment/natres/pdf/phosphorus/AgriPhosphorusReport%20final.pdf

44

4.4.2 Approach A

Micro-level assessment

Micro assessment of Water Abstraction (I.10), GNB (I.11-1) and Nitrates in Freshwater (I.11-2) using micro data at the agricultural holding level

The following section contains micro-level procedures for assessing net RDP effects for water abstraction in agriculture (I.10), GNB (I.11-1) and nitrates in freshwater, (I.11-2) from micro data collected through farm holding surveys. The procedure for assessing nitrates in freshwater (I.11-2) through a simulation 'case study' is explained immediately after this section.

Access to data and its quality and creation of consistent database and data infrastructure

I.10

The fiche for water abstraction in agriculture indicator points to two data sources for calculating the indicator. The 'Survey on Agricultural Production Methods' (SAPM), a one-off survey carried out in 2010 as part of the national censuses of agriculture and the Farm Structure Surveys (FSS),[49] recorded data at national and NUTS 2 levels. In this survey, water abstraction in agriculture, measured in cubic meters m^3 per farm per year and aggregated to the regional and national levels, is defined as the volume of water that has been used for irrigation on the holding during the 12 months prior to the reference date of the survey, regardless of the source.[50] A second source of data, is the OECD/Eurostat Joint Questionnaire on Inland Waters aggregated only on national territories. This database provides information on water abstraction for agricultural purposes and not actual use on the farm holding. As such, it takes account of the water losses in the storage and distribution network and of agricultural water uses beyond irrigation. From the OECD/ Eurostat data source, the WEI of Table 12 can be drawn. An additional source of data is 'Waterbase - Water Quantity' database that records water abstraction in agriculture at the River Basin District (RBD) and at national aggregate levels. The boundaries of RBDs do not coincide with the administrative NUTS boundaries. However, GIS methods and some simplifying assumptions can allow the calculation of the water abstraction index at regional level. The evaluator should be aware of the to use the same data source to construct a comparable time-series. This implies that the indicator should be re-calculated using historical data from only one data source. The above list of data sources is not exclusive. Different Member States have access to their own national data sources, that have in the past set up water monitoring networks for various purposes such as the monitoring of the Nitrates Directive, etc. Some of these networks became part of the European Environment Information and Observation Network (Eionet) and the WFD monitoring network.

I.11 - 1

The data available to calculate I.11-1 'GNB' are provided at national level by Eurostat.[51] At regional level there are no available data. Efforts to regionalise the national GNB data are still ongoing.[52] The estimation of GNB is based on an accounting identity of nutrient inputs minus nutrient outputs. For example, as concerns nitrogen budgets, inputs include mineral fertilisers, manure production and net manure import/export withdrawals, organic fertilisers, biological nitrogen fixation, atmospheric nitrogen depositions, seed and planting materials.[53] The most difficult input items to estimate at regional level

---

[49] Methodological details of the SAPM in each Member State can be retrieved in the methodological section of each reported census at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Agricultural_census_2010

[50] Part VIII of Annex III to Regulation 1200/2009 provides all irrigation related definitions.

[51] http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_gross_nitrogen_balance

[52] http://publications.jrc.ec.europa.eu/repository/bitstream/JRC96482/regnibal_final_pubsyjrc96482.pdf

[53] A detailed methodological account for the estimation of GNB is provided by the 'Methodology and Handbook Eurostat/OECD Nutrient Budgets' published in 2013 and accessible at: https://circabc.europa.eu/sd/a/4f405a13-9131-40c7-acea-b76f531da8b1/Nutrient%20Budgets%20Handbook%20(CPSA_AE_109)%20corrected3.pdf

are manure and fertilisers. Simulation models[54] at farm or at watershed spatial levels require this information to estimate GNB. Distributed simulation models such as the INTEGRATOR[55] may be viable alternatives for the estimation of I.11 GNB at regional level. Different Member States have access to their own national data sources for estimating GNB and/ or have developed methodologies to calculate GNB. For example, in the previous programming period in Poland, the amount of manure N was estimated from NUTS 5 stock densities and the amount of mineral N was calculated from crops grown according to the paying agency databases (ARiMR) and aggregated to a village level.[56]

I.11 - 2

Data for estimating the indicator I.11-2 'Nitrates' are available at the 'Waterbase-Water Quality' database maintained by the European Environment Agency (EEA).[57] This database contains national data delivered between 2000 and 2016 in the framework of the current Water Information System for Europe (WISE) State of Environment (SoE) - Water Quality (WISE-4) reporting obligation and River quality (EWN-1), Lake quality (EWN-2) and Groundwater quality (EWN-3) reporting obligations and cover the various indicators up to 2014. The spatial coverage is at the RBD or sub-unit levels[58] or the individual station.

At the micro level, water abstraction and GNB data are very sparse and fragmented. Certain Member States have access to water abstraction data through irrigation water registrars, water metering devices, or access to fertiliser purchases and actual consumption as a change of fertiliser stock. Some of these data are kept by farmers participating in various agri-environmental schemes. However, generally these data either do not exist, or are not very precise and/ or reliable. The best solution in such cases is to collect one's own data through a survey of farm holdings as it is explained below. A survey will also make the selection of counterfactual holdings easier and safer.

Selection of counterfactual option and micro-level method

For the assessment of I.10, I.11-1 and I.11-2 through a survey of agricultural holdings, the quantitative micro-level assessment follows the distinct steps developed below (Figure 11 in PART II, Chapter 2.6.2):

Step 1: *Obtain data for indicators I.10, I.11-1 and I.11-2* from national sources or produce an estimate of these indicators in case it does not exist (e.g. in the case of RDPs covering sub-national territory). This step is depicted in boxes 1, 2 and 3 of above mentioned Figure 11.

Step 2: *Retrieve data on common result Indicators R8, R12 and R13 (See the section 'explaining intervention logic')* and all available monitoring data that will reveal the number of supported agricultural

---

[54] Several countries have developed their own farm level nutrient management decision support systems or simulation models. The evaluator should ask the relevant authorities for the simulation software that has been found to be most suitable for the area under perspective. Good examples include PLANET developed by DEFRA and the Scottish Government in the UK (http://www.planet4farmers.co.uk/Content.aspx?name=PLANET), and NLEAP in the US (https://www.ars.usda.gov/research/software/download/?softwareid=428&modecode=30-12-30-15). INCA and EPIC are two examples of distributed simulation models which have been used in Europe. INCA is accessible at: http://www.reading.ac.uk/geographyandenvironmentalscience/research/inca/ges-inca-versions-applications.aspx and EPIC (Environmental Policy Impact Climate). For an EU wide application of EPIC see https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/water-requirements-irrigation-european-union and 'Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation.', Agricultural Systems 120, pp. 61–75, Balkovic et al., (2013).

[55] Accessible at: http://www.liaise-kit.eu/model/integrated-nitrogen-tool-across-europe-greenhouse-gases-and-ammonia-targeted-operational

[56] The case study in Poland is briefly presented in the 2010 Working Paper on 'Approaches for assessing the impacts of the Rural Development Programmes in the context of multiple intervening factors' (page 169). The document is accessible at: https://enrd.ec.europa.eu/sites/enrd/files/fms/pdf/EB43A527-C292-F36C-FC51-9EA5B47CEDAE.pdf

[57] 'Waterbase - Water Quality' provides a time series of nutrients, organic matter, hazardous substances and other chemicals in rivers, lakes and groundwater, as well as data on biological quality elements (BQEs) such as Phyto benthos and macroinvertebrates in rivers and lakes. Treated data at database (SQL) or spreadsheet (csv) format are available at: https://www.eea.europa.eu/data-and-maps/data/waterbase-water-quality#tab-european-data . Raw data and reports are available directly from Eionet at: http://cr.eionet.europa.eu/

[58] Waterbase stores information on more than 3500 river stations in 32 countries, more than 1500 lake stations, and quality data from around 1100 groundwater bodies.

holding population and the variety of measures used within the RDP's intervention logic. Use IACS/LPIS data to examine the spatial coverage of the supported agricultural holdings. Retrieve soil and land, use maps to examine the degree of environmental heterogeneity in the areas covered by the RDP's measures. This step is depicted in boxes 4, 5 and 6 of above mentioned Figure 11.

Step 3: *Decide if the number of supported agricultural holdings (from Step 2 above) is sufficient* for carrying out a proper evaluation either for water abstraction, or reduction of GNB and nitrates. Examine the possible heterogeneity of the RDP areas.

Step 4: Set up the counterfactuals in the following procedure:

Comparison groups will be created and the flow of operations from box 7 of above mentioned Figure 11 and following all the 'Yes' branches down to box 8 of Approach A.

a)   Decide on the group or sub-groups of supported holdings (from Step 2 above) and the control group, i.e. holdings which did not receive support from the RDP. At this step you will find out if there is a risk of double counting and how big it is.

b)   Decide on the spatial coverage of the survey as informed in Step 2 above. Decide on sample sizes depending on the method to be employed, the cost of sampling and the available budget.

F*or Nutrients in Freshwater (I.11-2) e*xamine if the WFD monitoring stations, or other national monitoring networks, address areas that are solely populated by supported or non-supported holdings, and examine their biophysical and agricultural heterogeneity. If such areas exist, then a survey of the holdings can provide, at least, a naïve statistics-based approach for netting out the effects on nitrates in freshwater. If such a situation does not exist, which is very likely, the assessment can be carried out using a simulation 'case study'.

Step 5: *Design a questionnaire* that will capture changes in water abstraction for irrigation (and for other uses) and changes in nutrients use. The questionnaire should include questions to capture the slippage effect, i.e. the fact that water conservation measures or nutrient reductions on part of the land, may release the same inputs to other parts of the holding and intensify production on the remaining land. Include also questions in order to capture any transverse effects, and especially the effects of M.01, M.02.1 and M.16.5, as well. You may decide to use on farm simulation models for extracting water abstraction and nutrients use. This will ensure objective and homogenous measurements. Examples of such simulation tools are presented in the box below.

Assessment of Farm Water Abstraction and GNB during the survey

During the field survey, the evaluator will collect data on water abstraction and GNB. Water abstraction data may be collected from metering devices if they exist, or from local water distribution authorities. As for fertilisers, some agricultural holdings keep records of bought and applied fertilisers by type. If such data are missing, or for cross-validating the data, the evaluator can use farm simulation models for water and nutrient requirements. The FAO's Aquacrop[59] is α model for estimating irrigation water needs for various crops and climatic zones that can be extended to work for whole regions with a GIS supplement. Several countries have developed their own farm level nutrient management decision support systems, or simulation models. The evaluator should ask the relevant authorities for the simulation software that has been found to be most suitable for the area under perspective. Good examples include PLANET developed by DEFRA and the Scottish Government in the UK,[60] and NLEAP[61] in the US. For example, NLEAP also incorporates a Nitrogen Trading Tool (NTT) analysis, which can be conducted to determine the potential benefits of implementing best management practices and the quantity of nitrogen savings that could potentially be traded in future air or water quality markets.

Net impact assessment at the micro-level: Once the data have been collected, coded and stored in a database, these can be analysed with an aim to estimate the net effect, scale up the estimations to the RDP level and verify the produced estimations with other sources by proceeding with the following steps.

Step 6: *Apply a method for analysing the data.* Regression and matching techniques are suggested depending on data quantity (sample size) and data quality:

a) Simple regression on water abstraction and GNB change with carefully chosen control variables that will reduce (but never eliminate) selection bias.

b) Instrumental Variables (IV) analysis deals better with selectivity, but is more demanding econometrically and requires the use of good instruments.

c) Construct a matching counterfactual from the sample of non-supported farm holdings with a matching algorithm. This procedure eliminates selection bias but requires larger sample sizes.

Step 7: In the rare case where *the MA has set up the survey before the start of the programme, or before and after data on water abstraction and GNB exist*, the methods in Step 6, and especially the matching algorithm, can be coupled with DiD methodology for more robust results as concerns sample selection (Box 9 of Figure 11, PART II, Chapter 2.6.1).

Step 8: *Estimate the Average Treatment Effect on the Treated* (ATT) to analyse the average effects the programme participation, on programme direct beneficiaries only, and compute the RDP's net direct effect coefficient for water abstraction and GNB.[62]

Step 9: If there are *indications of important indirect effects,* other than slippage (leverage) and transverse effects, either on supported or non-supported agricultural holdings, due to the application of water and nutrient reduction measures, which is rather unlikely, these should be treated separately. Qualitative methods and especially in-depth interviews with agronomists and local irrigation experts will allow the evaluator to highlight such possible effects.

---

[59] The Aquacrop website at: http://www.fao.org/land-water/databases-and-software/aquacrop/en/
[60] PLANET's site at: http://www.planet4farmers.co.uk/Content.aspx?name=PLANET
[61] NLEAP's site at: https://www.ars.usda.gov/research/software/download/?softwareid=428&modecode=30-12-30-15
[62] See Chapter 4.3.3.2 in Guidelines for Ex post evaluation of 2007-2013 RDPs

Step 10: *Aggregate the results and estimate the effects of the RDP at macro level.[63]*

Step 11: *Verify the above results with the qualitative data* obtained through interviewing experts, and by reviewing published case studies carried out on the RDP territory or in other RDPs with similar agricultural conditions.

Micro assessment of Nitrates in Freshwater (I.11-2) through a simulation 'case study'

It is often the case that, for assessing the net effects of the RDP on nitrates in freshwater, there are no monitoring stations connected only to participants or non-participants. In this case, the net effects of RDP on nitrates in freshwater cannot be assessed through a survey of agricultural holdings. The diversity of abiotic (physical, hydrological, climatic, soil, etc.), biotic (fauna and flora) and human factors (activities) that can be found within the boundaries of one RDP is enormous. This diversity can be captured by expensive and cumbersome methods such as the agricultural distributed simulation models (see box below). If such models exist for typical areas within the RDP's territory, then the MA can consider their application (Box 10 in Figure 11 of PART II Chapter 2.6.1). We suggest the evaluator chooses just one River Basin District (RBD) or RBD sub-unit that is 'typical' of the nitrogen pollution issues met within the RDP's boundaries. This must coincide with the area at which the survey of holdings was carried out for the purposes of Assessing I.10 and I.11-1. Then follow the generic procedure that is presented below:

Step 1: Choose a 'case study' area and a 'simulation' model. Several Member States have established monitoring programmes and/or simulation methodologies especially designed for nitrogen transportation, nitrites and nitrates.[64]

Step 2: Calibrate the simulation model with contemporary information on:

- Soils
- climate and weather
- land cover
- land use
- water and hydrography
- water abstraction by economic activities (e.g., manufacturing, power generation, etc.)
- nutrient deposition by economic activities (e.g., manufacturing, sewage, septic tanks, etc.)
- water abstraction
- nutrient use in agriculture

The last two can be the result of the farm holding survey described above.

---

[63] At this stage the evaluator has two options. If the survey is representative of the RDP territory, the evaluator can up scale the micro results for both supported and non-supported farm holdings. In the case where the survey is not representative of the whole territory, but only of specific targeted areas, the evaluator can work differently. One solution is to post-stratify the survey and assign weights to the cases to make them representative of the entire population. This is rather cumbersome, data and time demanding. Alternatively, apply the net direct effects coefficients on the total RDP effects as these are captured by the relevant result indicator and estimate only the total RDP net direct effect. Compare the total water and GNB reductions as estimated by I.10 and I.11-1 with the RDP's net direct effects. Does the RDP have an important contribution?

[64] For example, the UK has, since 2004, established 16 study micro-catchments representative of a range of environmental and agricultural conditions. Farm practice and water quality monitoring data were collected and used to validate N loss models. In Germany, Kreins et al. (2009) developed an interdisciplinary model network consisting of the regionalised agricultural and environmental information system RAUMIS, the hydrogeological model GROWA/WEKU and nutrient emissions in river systems MONERIS to analyse the impacts of nutrient reduction measures on the water quality of a 49,000 km^2 catchment. In Greece, INCA-N and INCA-P simulation models were used to assess the efficiency of agri-environmental programmes in reducing N and P nutrients. Works available at:
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.511.4712&rep=rep1&type=pdf and
https://www.researchgate.net/publication/309732435_False_Positive_and_False_Negative_Errors_in_the_Design_and_impl
ementation_of_Agri-environmental_Policies_A_Case_Study_on_Water_Quality_and_Agricultural_Nutrients

Net Impact Assessment

Step 3: For the evaluator, the calibration results are the outcome of the prevailing water abstraction and fertiliser application rates at the time the model will be calibrated. This therefore includes the RDP effects and provides an estimate of the situation 'after'.

Step 4: The evaluator can simulate the results before the operation of the RDP by increasing the application and abstraction rates by as much as is shown by the survey of agricultural holdings for water abstraction and GNB, for supported and non-supported holdings. This is assuming everything else to remain the same or change the baseline according to information on changes in other activities (e.g., connection of a village to a sewage system, etc.). This will provide an estimate of the situation 'before'.

Step 5: Net impact assessment is the comparison between the situation 'before' and 'after' having taken into account all changes. The 'before' and 'after' change can be estimated for nitrates concentration in specified monitoring stations (micro assessment) and/or for all monitoring stations together (watershed level assessment), depending on the detailed application of the simulation model.

Step 6 (optional): Test if the applied agri-environmental measures are 'climate change proof' by simulating the baseline on different weather and hydrological data according to climate change projections.

Distributed Simulation Models

There is a wealth of distributed simulation models for water, sediment and nutrient transportation such as EPIC,[65] FAO's CROPWAT,[66] the WASIM,[67] the SWAT[68] and the INCA family of models.[69] Many of these models can be combined with pure hydrological models, such as the semi-distributed PERSIST,[70] and the well-known climate change models (e.g., HadCM3 and ECHAM) or other nutrient and sediment transportation models. These comparisons will allow the production of estimates of Eflows under climate change scenarios, or link water abstraction with water quality and soil erosion. The model INTEGRATOR is an N budget distributed model linked also to Green House Gasses.[71]

Macro-level assessment

Access to data and its quality & creation of consistent databases and data infrastructure

Macro level data vary enormously among Member States and RDPs. Macro data may exist at different spatial levels or may not exist at all. Certain Member States maintain detailed environmental databases that are spatially disaggregated to the NUTS 3 or lower levels. Such data refer to water abstraction, GNB, or good proxies such as fertiliser consumption by type of nutrient, etc. Availability and access to macro level data should be reviewed on a case-by-case basis.

---

[65] The model's acronym now stands for Environmental Policy Impact Climate. EPIC is well suited to calculate irrigation water requirements by setting constraints to timing and irrigation rates.
[66] Accessible at: http://www.fao.org/land-water/databases-and-software/cropwat/en/
[67] Accessible at: http://www.wasim.ch/en/
[68] Accessible at: http://swat.tamu.edu/
[69] INCA's web page at: http://www.reading.ac.uk/geographyandenvironmentalscience/research/inca/ges-inca-versions-applications.aspx . INCA-N is INCA's is an example of a distributed nitrogen transportation model with many applications in Europe. INCA-P is the corresponding model for phosphorous transportation.
[70] PERSiST's site at: http://www.slu.se/en/Collaborative-Centres-and-Projects/slu-water-hub/models/persist/
[71] Accessible at: http://www.liaise-kit.eu/model/integrated-nitrogen-tool-across-europe-greenhouse-gases-and-ammonia-targeted-operational

Selection of a counterfactual option and macro-level method

The macro level approaches, due to various sources of existing information or, due to non-availability of data should be judged by the evaluator on a case-by-case approach. Below four different methods are reviewed.

1. For those MAs that can use strong environmental databases (such as the German RAUMIS) at lower spatial level, i.e. the NUTS 3 level or lower, for water abstraction and GNB, the use of a **Generalised Propensity Scoring Matching (GPSM)** and/or **spatial econometrics methods** coupled with DiD are the most suitable methods for netting out the RDP's effects in a macro assessment framework.

2. In case of I.10 - water abstraction, data at the NUTS 3 level may be compiled using the **GIS methods** (overlay River Basin District boundaries with administrative boundaries) and by adopting some simplifying hypotheses to produce matching data on water abstraction and support for each NUTS 3 area. Data may also exist from irrigation agencies, monitoring irrigation water. Even better, if support data are georeferenced and can be aggregated at the RDB level, the evaluator will not have to adopt any simplifying assumption and will produce the matching support-water abstraction data at the RBD or RBD sub unit level.

3. If for the above method the evaluator does not have access to an environmental database, or the RBD units may not be enough to justify a robust statistical methodology, the methods may be restricted to **simple regressions** (without controls) or **naïve group comparisons**. For I.11-2, if the monitoring stations are numerous within RBDs, the same methodology may be applied, i.e., percentage of monitoring stations in a certain quality class within an RBD.

4. **Distributive agricultural models** such as the EPIC, which has been used for estimating irrigation water requirements in Europe,[72] can be used to assess the I.10 water abstraction in agriculture. The model can be calibrated for the land cover and agricultural land uses at the start of the programme and simulate the RDP's effects, resultant from reduced water abstraction as informed by the micro assessment. EPIC's major pre-requisite is the existence of a map of irrigated areas. Of course, other simulation models exist and many academic and research institutions around Europe have experience with simulation models and have already calibrated various models for specific physical- and bio-geographic areas.

Micro-macro consistency and validation

For the net impact assessment, it is important that the results of both micro and macro level assessment are consistent, i.e. results of these assessments show the same trend in relation to impact (water abstraction or GNB), even if the evaluator used a proxy indicator in case the impact indicator could not be estimated (e.g., fertiliser consumption instead of GNB at regional-macro level), or even if different methods for the assessments were employed. Triangulation of net impacts with information from qualitative sources, published case studies or academic work should be performed. For nitrates in ground or surface water monitoring stations assessed by a ''case study'' approach that uses the micro water abstraction and GNB assessments for calibrating the model and simulating scenarios, there is no scope for checking the consistency between micro and macro level estimates. However, there is scope for verifying that the chosen watershed is a 'typical' watershed that allows the evaluator to draw conclusions valid at a national level.

---

[72] Water Requirements for Irrigation in the European Union at: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/water-requirements-irrigation-european-union and 'Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation', Agricultural Systems 120, pp. 61–75, Balkovic et al., (2013).

Pros and cons for using Approach A

Important feature	Pros	Cons
Simple regression methods do not demand high econometric skills, but their ability to reduce selection bias and produce unbiased net effects estimates depends on careful design of the survey and collection of 'good' control variables.	Easiness of application.	Possibility of bias is not eliminated.
Instrumental Variables (IV) approach is econometrically more demanding than a simple regression but, if there are suitable instruments, it deals well with selection bias.	Selection bias is adequately dealt with.	Intermediate econometric skills are required.
Matching algorithms and in particular the Propensity Score Matching, is the best method for dealing with selectivity bias but it requires a larger sample size for non-RDP participants in order to ensure that the largest proportion of RDP participants can be matched to a non-participant.	Best method for addressing selection bias.	Larger sample sizes and intermediate econometric skills are required.
Simulation models for nitrates in freshwater can 'isolate' the effects of agriculture out of a very complex network of sources contributing nutrients to the watercourses.	The impact of agriculture on nitrates concentration in freshwater is very well approximated. Various scenarios can be tested, including climate-change scenarios.	Very specialised and demanding methodology requiring a lot of data and advanced hydrological modelling skills.

4.4.3 Approach B

Approach B (Box 13 in Figure 11, PART II, Chapter 2.6.1) should be used if comparison groups do not exist or cannot be set up (Box 7 in above mentioned Figure 11), a sufficiently accurate model does not exist (Box 11 in Figure 12) and there is no appropriate time for evaluation (Box 12 in above mentioned Figure 11). Approach B (Box 13 in above mentioned Figure 11) also can be used if the RDP's uptake of the relevant measures is too low to justify the cost of a survey. In such cases there are two alternatives:

Naïve Group Comparisons. This alternative compares an average of the change in impact indicators in supported farm holdings with a grand average that is considered to be the change in the impact indicator in the absence of the RDP. The work can proceed in the following steps:

Step 1: The average of the change in the impact indicator for supported holdings is constructed. This can be done from monitoring data including application forms of RDP supported holdings. For example, the complementary indicator R13 can provide an estimate of water abstraction reductions for supported

holdings. Agri-environmental plans or applications for the adoption of agri-environmental schemes can provide data for reductions in applied nutrients or water abstraction reductions.

Step 2: The 'counterfactual' is the corresponding average of the NUTS 2 area or other wider area in which supported holdings are located. For water abstraction, the wider area (NUTS 2 or water department, etc) average may be constructed by using WFD data and IACS/LPIS data. The bio-physical conditions and agricultural patterns of the area must resemble the corresponding characteristics of the area where the supported agricultural holdings are located. For GNB it may be more difficult to construct population averages for certain geographic areas. This may be facilitated by environmental databases in the Member States that have access to them (see above) or by using national averages.

Step 3: Estimate a 'net' effect by comparing the average of the 'participants' from Step 1 above to the counterfactual from Step 2 above.

Step 4: If from monitoring data the evaluator can calculate a before (application forms) and after water abstraction level or nutrient field deposition level, then a naïve DiD approach can be adopted. This procedure should be complemented by a sensitivity analysis of the outcomes and qualitative data collected from in-depth interviews with soil experts and agronomists, and possibly a focus group that will shed light on the magnitude of indirect effects such as a possible slippage effect or multiplier effects.

Qualitative assessment: This alternative applies only qualitative methodologies in the form of focus groups, Delphi or the MAPP method. Sources for framing non-supported agricultural holdings may be acquired by contacting irrigation water authorities (who have installed water and pumping metering devices), electricity consumption authorities, fertiliser distributors, agronomists or other specialists. This information can be useful in deducing water use and fertiliser application coefficients for non-supported agricultural holdings for the same areas and, if possible, the same Type of Farms (TFs) where support was provided. This can provide an indication of the trend in the use of water and fertilisers, an indication of the deadweight and of the leverage whilst providing a crude indication of other secondary effects. Finally, information may be collected for the role of supporting measures such as M01 and M02.1 or M16.5 by instructing advisors and training authorities or by advising possible innovation and cooperation partners to flag possible transverse effects.

Pros and cons in using Approach B

Important feature	Pros	Cons
Naïve group comparison methods.	Easiness of application, especially if complementary results indicators exist or can be calculated from applications data.	Net effects are only approximate. Selection bias is not addressed.

Preconditions for applying Approach A and B

> **Survey**
>
> The MA´s decision whether a survey will be applied in the evaluation of RDP effects on water abstraction and quality depends on the number of supported farm holdings. It is always useful to conduct the survey which collects data for all environmental indicators (The open issues to be clarified have been discussed in similar boxes for GHG and ammonia emissions in Chapter 2.3.1, of these Guidelines). Specific questions to be answered are:
>
> - Will it be an RDP territory-wide survey or will it target specific 'problematic' RDP areas? Problematic areas may refer only to WFD areas with acute water quantity issues, or may address Nitrate Vulnerable Areas, or both.
> - Will it address a wide range of measures or will it be restricted to only a few (2-3) measures? If there are many (and overlapping measures) to examine how to deal with the risk of double-counting?
> - Will you use a water and nutrients calculator for estimating that water abstraction and nutrient application change uniformly with one method?
>   Simulation model:
> In case the MA wants to apply a simulation model in the assessment, it must ask the evaluator:
> - To prove experience with simulation models in the same area or
> - To use an already operating simulation model in the RDP area, if it exists (which is safer and cheaper).

4.4.4 Adequateness of suggested evaluation approaches

The above described evaluation approaches are in the following discussed as regards their adequateness in fulfilling the evaluation quality criteria: rigour, reliability, robustness, validity, transparency, credibility, practicability and cost effectiveness. The definitions of the quality criteria are provided in Table 2.

Table 9. Adequateness of the proposed evaluation approaches for the assessment of CAP common impact indicators I.10 – Water abstraction in agriculture and I.11 Water quality

Quality Criteria	Approach A (optimal)	Approach B (alternative)
	Micro-level **Regression and matching techniques for I.10 and I.11 - GNB and for I.11 - nitrates in fresh waters in case of data availability** Macro approach **GPSM, Spatial econometric methods**	Micro- and macro-level **Naïve group comparisons supported by qualitative methods**
Rigour	**Approach A is rigorous** At micro-level the econometric methods can produce robust results depending on the methodology used. For macro studies, if the sample size is adequate and/or if coupled by a spatial environmental database, the evaluation findings are rigorous	**Approach B is partly rigorous**. Naïve comparisons between supported and average population can only show the trends for changes and can produce a biased estimate. However, bias may not be that serious depending on the number of supported holdings out of total number of holdings in the region. Also, qualitative methods can only show the trends and can produce a very subjective estimate.
Reliability	**Approach A is reliable** at micro-level, if the sample size is adequate. At macro-level the estimates will depend on the harmonisation of simulated data and good micro assessments.	**Approach B is partly reliable.** At micro-level the reliability depends on the size of the selection bias that is introduced by naïve group comparisons.

Quality Criteria	Approach A (optimal)	Approach B (alternative)
	Micro-level **Regression and matching techniques for I.10 and I.11 - GNB and for I.11 - nitrates in fresh waters in case of data availability** Macro approach **GPSM, Spatial econometric methods**	Micro- and macro-level **Naïve group comparisons supported by qualitative methods**
		At macro-level the approach is able to produce stable and consistent indications of the trends, but not exact figures.
Robustness	**Approach A is robust.** At micro-level the estimates depend on the choice and availability of variables to control the matching algorithm or the regression. At macro-level the robustness is ensured with the adequate number of special units.	**Approach B is partly robust.** Its robustness depends on the size of the population of supported holdings. Using qualitative assessment can produce stable and resilient findings but only to show the trends of the impact.
Validity	**Approach A is valid.** At micro-level, statistical estimates of errors are produced for micro studies and should always be taken into account when findings are the subject of a sensitivity analysis. For the simulated models the approach is also valid, however, they should always be cross-validated.	**Approach B is partly valid,** depending on the number of cases, statistical estimates of errors may not be possible to be estimated or their confidence intervals may be very broad.
Transparency	**Approach A is transparent.** All data and methods are publicly available, and all processes are recorded and unified.	**Approach B is transparent.** All data and methods are publicly available, and all processes are recorded and unified.
Credibility	**Approach A is credible.** Selection bias is well handled by the proposed econometric methods. However, the rapid changes of the institutional framework, and especially of the ex-ante conditionalities for water, must be taken into consideration when judging the credibility of the approach.	**Approach B is partly credible**, because selection bias will be present. The severity of the bias depends on several factors. However, qualitative information can point to areas where selection bias may be an issue.
Practicability	**Approach A is practical.** Yes, the methods and estimation procedures are fully practicable but difficult and cumbersome.	**Approach B is practical.** The methods and estimation procedures suggested by this approach are fully practicable.
Cost-effectiveness	**Approach A can be costly.** This depends on the cost per sampled unit and the sample size. The ability to pre-process data for RDP supported holdings will reduce the cost.	**Approach A is cost effective.** It can provide a cost-effective way of assessing water abstraction, GNB and nitrates in freshwater impacts under limited data availability.

4.4.5   Dos and don'ts

Dos

- Understand the institutional nexus underlying irrigation water abstraction, and the changes induced by the WFD and RDP's ex-ante conditions.

- Contact all institutions that calculate or estimate regional and national water abstraction alongside water quality figures reported in Eurostat or the WFD. Especially the WFD focal points.

- Ask for the latest available figures and/or additional indicators even if they are unpublished. Establish a time line that includes the RDP's time frame.

- Locate institutional users of irrigation water simulation models, or of N and P budget models (distributed or farm specific), that are already in operation and calibrated.

- Explore whether you can develop synergies with other evaluations that may have the same approach, e.g., evaluation of soil erosion and/or soil organic matter before starting to build the database.

- Review your monitoring database and application records to locate data gaps.

- Coordinate the possible application of a simulation model with the GNB field survey.

- Prepare and pre-process as much data as possible for beneficiaries, from applications and agri-environment plans.

- Search for existing farm holding sampling frames either specifically addressing water abstraction (e.g. SAPM) or water quality (WFD monitoring stations), or more general sampling frames (e.g., FADN, IACS Payment Authorities, etc.).

- Search for existing GIS maps with irrigated plots and soil types or georeferenced IACS.

- Clarify the criteria (eligibility and locational) that would categorise a farm holding as the RDP beneficiary.

- Decide on your sample size, questionnaire structure and good control variables (observables) as early as possible.

Don'ts

- Mix data sources when producing a time line for the I.10 or other indicators, in order to ensure consistency and coherence.

- Construct a sampling frame from scratch with the support of FADN, water registered users, local cooperatives when sampling frames for non-beneficiaries are missing, etc. Sampling non-beneficiaries is essential to your work.

4.5 CAP common impact indicator I.12

The following sections of the Technical Annex are related to Chapter 2.7 in PART II of the Guidelines.

4.5.1 Additional indicators (examples)

Additional indicators which could potentially be used to assess the RDP effects on solid organic matters in arable land are described below in detail:

1. SOC 0-60 cm: the impact indicator I.12 assesses the total SOC in arable land solely in the topsoil layer (0-20 cm). This provides partial (incomplete) information concerning the content of soil carbon in arable fields since only about one-third of total SOC is located in the upper 30 cm of the soil[73] [74] (and probably even less in the top 20 cm – the depth suggested by the impact indicator I.12). In order to obtain more comprehensive information about the SOC in arable fields, arable soil should be sampled up to 60 cm instead of 20 cm.

2. SOC change: Due to carbon-unfriendly cultivation practices, most arable soils in the EU-28 have lost a portion of their SOC. Currently their storage capacity is not fully used, and they have not reached their carbon sequestration saturation point. The additional indicator 'SOC change' shows how much carbon has been stored or lost in arable land compared to the previous reporting period, as well as the rate of this change. The indicator does not require any additional soil sampling or soil analysis. It can be calculated (extrapolated) from the existing data for I.12 (soil depth 0-20 cm) or data for the additional indicator SOM 0-60 cm.

3. SOM bio: the content of SOM in soil is pretty stable. Short-term changes, including potential increases due to the implementation of RDP measures occur gradually and are not easily detectable due to the high background of soil C level. In most cases it takes a decade or so before a significant change is achieved. Since the RDP agri-environment contracts only last for five years, it is of the utmost importance to identify reliable short-term indicators for organic matter changes. We propose to include an additional indicator which would assess biologically-active forms of SOM that function as short-term indicators of longer-term changes in SOM. This approach is in line with EIP-AGRI[75]suggestions and the fiche on the common impact indicator I.12 Soil organic matter in arable land, which states that the I.12 indicator 'should be ideally complemented by an assessment of soil biodiversity'. The 'SOC bio' is calculated as a composite index comprising the following three indices of similar weight:

    o Microbial biomass C (C_{mic}) to total organic C (C_{org}) ratio (C_{mic}/C_{org});

    o Microbial respiration (CO_2) to microbial biomass C (C_{mic}) ratio (CO_2/C_{mic});

    o ßeta-glucosidase enzyme activity (ßeta-glucosidase units/L) to microbial biomass C (C_{mic}) ratio (ßeta-glucosidase units/L /C_{mic}). The ßeta-glucosidase enzyme is preferred over other enzymes because it integrates information on microbial status and soil physical-chemical conditions in a satisfactory manner.[76]

So, the 'SOM' bio indicator should be calculated using the following formula:
SOM bio = (C_{mic}/C_{org}) x (CO_2/C_{mic}) x (ßeta-glucosidase units/L /C_{mic})

The soil sampling for all three microbial tests should also be done at a 0-60 cm depth. However, in cases where soil sampling for the purpose of the assessment of I.12 is taken only at the depth of 0-20 cm, the SOM bio assessment can also be done at such a depth.

---

[73] 'The Vertical Distribution of Soil Organic Carbon and Its Relation to Climate and Vegetation.', Ecological Applications, 10, 423–436, Jobbágy E. B., Jackson R., (2000).

[74] 'Distribution of Organic Carbon in Soil Profile Data, Joint Research Centre', Ispra, Hiederer R., (2009).

[75] 'Mini paper: Data Collecting Issues and Measuring. EIP-AGRI, Brussels.', Costantini E., Sideris C., Dell'Abate M.T., Mavridis A., Anastasiou D., (2016).

[76] 'Temporal and spatial evolution of enzymatic activities and physic-chemical properties in an agricultural soil.', Applied Soil Ecology 18: p. 255, Aon M.A. and Colaneri A.C., (2001).

The overview of additional indicators to measure soil organic matters in arable land, their measurement units, the information on their use and data sources/frequency is in the Table 11.

Table 10. **Examples of additional indicators to assess SOM in arable land**

Indicator	Measurement unit	Use of additional indicator	Data sources and frequency of collection
SOC 0-60 cm	1. Total SOC in arable land in megatons; 2. Mean SOC concentration in arable land in g/kg.	Provides information about the content of SOM in the layer up to 60 cm. This is considerably more reliable than results obtained from I.12 based on the top-layer of up to 20 cm.	Soil monitoring programme database. Once to four times a year.[77]
SOC change	1. Total SOC in arable land in megatons (calculated as the difference between the previous and the current reporting period); 2. SOC change in percentage (calculated as the difference between the previous and the current reporting period).	Provides information about whether compared to the previous reporting period SOC is increasing or declining, and at what rate.	In case a soil monitoring programme database is available: annually (at the time of reporting). In case LUCAS database[78] is used: once in three years.
SOM bio	Index of biologically-active forms of SOM (for details see the text above the table).	Provides indication of short-term SOC changes, since the proposed common impact indicator I.12 is a poor indicator for detecting gradual SOC changes.	Soil monitoring programme database. Once to four times a year.

---

[77] Frequency of soil sampling differs among Member States/RDPs. Once of year is considered as a minimum. In case soil sampling is done more than once a year, the annual value of both total SOC and mean SOC concentration is calculated as the mean of all samplings. It is strongly advised to have more than one sampling per year because of temporary variations in SOC.

[78] Land Use / Land Cover Area Frame Survey (LUCAS), available at: https://esdac.jrc.ec.europa.eu/projects/lucas

4.5.2 Approach A

Micro level assessment

Access to data and its quality & creation of a consistent database and data infrastructure

For the optimal assessment of SOM in arable land it is of the utmost importance to have access to a **sound field monitoring programme** comprising a robust, high-quality soil sampling and analysis, both at beneficiary and non-beneficiary sites. This is because comprehensive data on SOM at macro level (regional and national) are scarce in most EU Member States. Down-scaling from regional or national soil monitoring programmes to micro level (e.g. field) is very difficult because most of these have insufficient soil sampling density, leading to methodologically questionable, unreliable and partial results. Implementation of field monitoring programmes can be demanding and cumbersome. They require on-field soil sampling, laboratory analysis and expert skills, also tending to be time-consuming and relatively expensive. Implementation of soil monitoring programmes requires good designing and implementation arrangements. In practice, it can be challenging also due to the large spatial and temporal variability of soil carbon, sampling methods and sampling frequency; heterogeneity of measurement techniques and interpretation of results. **Especially considering that at the EU level no uniform soil sampling and analytical (laboratory) protocols for the assessment of SOM have been defined and adopted.**

Moreover, all commonly used existing methods and practices have shortcomings and there is no any preferred or 'best' method. Relying solely on the existing soil monitoring programmes is a cost-effective approach but most of these programmes are not likely to be able to provide appropriate data for the assessment of I.12. Most of them can hardly produce sufficient information for creating a good GIS which would be able to overlay with LPIS data to have a counterfactual.

> **SOC analysis in situ**
>
> At this stageplace, it would also be fair also to mention that SOC analysis can be done also *in-situ* – in the field by means of portable spectroscopy. This is a cost-effective and non-destructive technique allowing for the assessment of a much larger number of sampling locations compared to that offered by lab-based methods. However, the accuracy of this method is low. Vegetation or crop residue cover, differences in soil surface roughness and soil moisture can interfere with and disturb the spectral signal, influencing soil analysis results. Since this method cannot achieve an acceptable level of accuracy – it cannot be advised. The same goes for remote spectroscopy – the use of reflectance spectroscopy on airborne (e.g. drones, airplanes, etc.) or satellite-mounted sensors, employing Near and Mid Infrared Reflectance Spectroscopy (NMIRS). This is a cost-effective method which can provide a rapid soil appraisal at a larger scale. However, it is not able to detect vertical gradients in SOC within the topsoil.

Managing Authorities are therefore strongly advised and encouraged to set-up and implement sound soil monitoring programmes with the robust data sampling density and frequency. Without these they will not be able to properly assess I.12. Soil monitoring programmes can be used jointly for the assessment of I.12 and I.13 (soil erosion). Their implementation can be financed from M20: Technical Assistance. In some cases, MA might consider linking the soil monitoring programme for SOM (and erosion) with the on-going soil monitoring programmes set for scientific or other purposes.

Selection of a counterfactual option and micro-level method

The following steps are proposed for the assessment of I.12:

Step 1: *Build comparison groups:* Involve both RDP beneficiaries and a comparison group comprised of non-beneficiaries in the sound soil monitoring programme.

Step 2: *Obtain soil samples:* Soil samples should be taken at two depths: 0-20 cm and 0-60 cm. The samples from the depth of 0-20 cm allow for the assessment of indicator I.20 – as requested in the fiche, while the samples taken at the depth of up to 60 cm will enable analysis and application of all three additional indicators throughout the soil root zone of most arable crops. The soil samples can be taken following instructions described in the most recent LUCAS publication.[79] It is advised to have more than one sampling per year because of temporary variations in SOC (the soil microbial community is particularly sensitive to drying and rewetting and these effects can last for more than a month after the last stress).[80] For the laboratory analysis we recommend using the method of determination of organic and total carbon after dry combustion (elementary analysis) – ISO 10694:1995.[81] This is the same method used by LUCAS. It is recommended to record the vegetation type and date of sampling, as well as laboratory methods of testing. This is because no standardised methods for soil testing exist, so this information might help when comparing results across the RDPs.

Step 3: *Up-scale at RDP level:* Up- scale the data obtained from the soil monitoring programme to RDP level. A database should be created enabling an easy access and overview of results both for the beneficiaries and the comparison group, as well as aggregation of data at RDP level by summing-up/aggregating all individual data from both groups. The results should also be used to build a digitalised soil map (1:10,000), linked with GIS database, preferably also containing information on soil type and soil texture. This should enable to compare changes in SOC in time and space scale.

Net impact assessment at micro level

Step 4: *Assess net value of impact indicator I.12* via comparing changes in SOC on the time and space scale between beneficiaries and non-beneficiaries, through extrapolation of data from the above-mentioned database.

The expected difference between the gross and the net effect for I.12 is likely to be negligible because the SOM changes gradually and because there are no major external factors (e.g. climate, market conditions, etc.) influencing the dynamics. In addition, most soil carbon conservation techniques are expensive. They require sophisticated and expensive machinery (e.g. for reduced tillage), large amounts of external carbon-rich sources (e.g. livestock manure, food-industry by-products, etc.) and change in management practices, such as the change in crop rotation comprising a higher share of soil carbon enhancing crops, like grass-clover, green manures, etc. This results to additional costs and/or foregone – at least for short and mid-term. It is therefore quite unlikely that many farmers will be able to implement the above without benefiting from the support for RDP measures, contributing to carbon conservation in the soil. In some cases, there may be a market incentive leading to the implementation of these practices – for instance cultivation of grass-clover because of a high market demand and good price. In that case, more farmers will grow grass-clover and potentially increase soil carbon stock on their arable land. However, in practice, it rarely happens that market incentive initiates a shift to practicing soil carbon conservation agri-methods.

Macro level assessment

There are no macro level assessments that could be employed to realise the optimal evaluation approach. It is not possible to downscale any of the existing soil monitoring databases because none of them contain results of soil analysis based on soil sample to the depth of 0-60 cm.

---

[79] 'Soil component: Sampling Instructions for Surveyors.', EUR 28501 EN, doi 10.2760/023673 LUCAS, Fernández-Ugalde O., Orgiazzi A., Jones A., Lugato E., Panagos P., (2018).

[80] 'Effects of drying and rewetting frequency on soil carbon and nitrogen transformations Soil Biology & Biochemistry.', 34, 777-787, Frier, N. and Schimel, J.P., (2002).

[81] See at: https://www.iso.org/standard/18782.html

Micro-macro level consistency and validation

Since Approach A does not envisage any macro level assessment there is no need to undertake steps ensuring micro-macro level consistency and validation.

4.5.3 Approach B

Micro level assessment

Access to data and its quality & creation of a consistent database and data infrastructure

As already shown in Figure 13 of Chapter 2.7.2 in PART II, the alternative approach has three levels of: high, medium and minimum-level assessment. The first two are entirely based on a micro level assessment – soil monitoring programmes comprising soil sampling at either 0-20 cm depth or both 0-20 and 0-60 cm depth. The high-level assessment is identical to the approach A, except that it does not provide data for the use of SOM bio indicator. The medium level assessment relies on the assessment based on a robust soil monitoring programme at the depth of 0-20 cm, providing data for the use of indicator I.12 and two additional indicators: SOC change and SOM bio. The minimum level assessment can be based on one of the three paths. Path 1 is the only one solely relying on micro level assessment (soil monitoring) (above mentioned Figure 13).

Selection of a counterfactual option and micro-level method

Recommendations on soil sampling and laboratory analysis presented under Approach A are also valid for the micro level assessment of Approach B.

Net impact assessment at micro level

The net effect can be assessed by using qualitative methods such as organising focus groups, structured interviews or by conducting surveys among farmers and extension officers. These can provide valuable information from which the net effect can be assessed. If needed, a regression analysis and regression coefficients may be employed in order to provide more insight into the matter.

Macro level assessment

Access to data and its quality & creation of a consistent database and data infrastructure

The **macro level assessment is applicable only in cases where no sound soil monitoring programmes exist.** It leads either to Path 2 or Path 3 in the above-mentioned Figure 13.

Path 2 is a combination of soil monitoring and LUCAS data, while Path 3 solely relies on Figures derived from LUCAS data, complemented by data from other sources. In Path 3, the evaluation of indicator I.12 both for the beneficiary and comparison group has to rely on LUCAS data. However, this approach faces significant challenges and provides weak results. The robustness of the LUCAS dataset is not sufficient. The current version contains 19,969 topsoil samples (0-20 cm) from 25 Member States (data for Bulgaria, Romania and Croatia are missing), taken in 2009. Of this, only 7,601 soil samples (38 per cent) originate from arable land, and the density of sampling sites is insufficient to assess I.12. In most Member States, it is less than one soil sample per 10,000 ha (100 square kilometres) of arable land. In only five Member States, the density is higher than 1 soil sample per 10,000 ha of arable land, while in five Member States it is lower than 0.5. The LUCAS database should be scaled down and eventually overlaid with LPIS data to obtain the counterfactual. However, given the small density of data in LUCAS,

downscaling data on NUTS 2, NUTS 3 or agricultural holding level will be a challenge in most Member States.[82]

The average per hectare value of SOC in arable land for each Member State can be obtained if the total estimate of organic carbon content in arable land[83] is divided by the total number of hectares of arable land. However, such a figure does not tell the whole story and is lacking the practical use. Spatial, temporary, climatic, geographical and paedogenetic arable land variations exist regionally, and even locally within each Member State. Thus, a national average value of SOC content in arable land would have no practical value for the purpose of evaluation of indicator I.12.

Information on databases on SOM on agricultural land in general, and arable land in particular, at Member State level including their regions, is scarce. France, Italy and Spain have large data sets on SOM content in soils, although not properly organised, whereas most other Mediterranean countries of Europe have only limited data, or data from field surveys that are either insufficiently georeferenced or not accessible outside the country of origin categories.[84]Scales of (systematic and non-systematic) sampling schemes differ greatly across Member States: from 1: 5,000 in Greece to 1: 1,000,000 in France and Belgium (Flanders). [85]

Complementary information for assessment at macro level

Because of the problems mentioned above, downscaling of LUCAS data in Path 3 of the minimum level assessment must be **complemented with information from supplementary sources** (above mentioned Figure 13). In many countries/regions**, independent soil monitoring programmes exist**. These are usually set-up to serve scientific research. Although such soil monitoring programmes might not fully fulfil requirements set by RDP evaluation of I.12, their results can be a useful auxiliary tool that should complement LUCAS data. The same goes for various SOM models that might exist at research institutes and Universities, as well as results of demonstration trials, small-scale experiments, PhD. research and scientific publications. All these should be employed to complement LUCAS data. However, if such supplementary sources of information are not available at all (which is very unlikely), the evaluators should **complement the LUCAS data with qualitative analysis** using Focus Groups, Delphi, MAPP, etc. However, both the MA and evaluators should be aware that the use of qualitative analytical methods for the assessment of I.12 will be exceptionally allowed only for reporting in 2019. Qualitative methods do not provide hard data and are not the way out of the problem. The indicator I.12 can only properly addressed through sound soil monitoring programmes.

Micro-macro level consistency and validation

The micro-macro level consistency and data validation is only relevant for Path 2 of the minimum level acceptance, because Path 2 combines soil monitoring and LUCAS data. The number of LUCAS soil samples at the level of individual RDPs will be relatively small, which facilitates an easier consistency and validation check. The evaluators should check if data on SOC from individual sampling points are within the usual value ranges (usually in the range of 0.5 to 4 per cent). Values outside this range should be double-checked. Since RDP soil monitoring programmes might be based on different soil sampling and laboratory analysis techniques than used in LUCAS, comparison and data validation between the two datasets is often not possible.

---

[82] Besides LUCAS, there seems to be no other relevant databases containing information on SOC. Data on SOC content in arable land cannot be found in, or extrapolated from, any of the following databases: FADN; EU Farm structure surveys; Paying Agency databases; Databases of the Ministries of agriculture and environment; the FAO statistics.

[83] Available from the EC's Web page on the CAP context indicators, see C. 41, available at: https://ec.europa.eu/agriculture/cap-indicators/context/2017_en
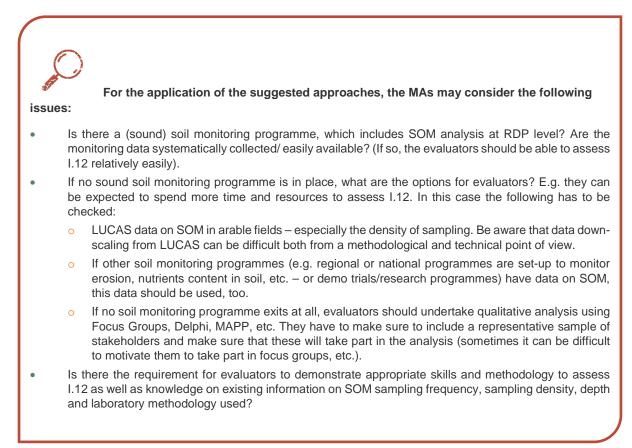
[84] EIP-AGRI, 2015. Focus Group Soil Organic Matter in Mediterranean regions, Brussels.

[85] 'Current status of risk assessment methodologies for soil organic matter decline in Europe.', Joint Research Centre, Ispra, Tóth G., Kuikman P., Ehlert P., Chardon W. and van Beek C., (2009).

Pros and cons in using Approach A and B

Important feature	Pros	Cons
Evaluation of I.12 based on data from a sound field monitoring, designed for the purpose of evaluation of RDP measures; implemented at RDP level and comprising a robust, high-quality soil sampling and analysis both at beneficiary and non-beneficiary sites.	Provides most reliable results. Can be multi-purpose, combined with monitoring of other RDP soil and water-related indicators. Once established and put in operation it is relatively easy to maintain it.	Demanding: it involves on-field soil sampling, laboratory analysis and expert skills, and tends to be time-consuming and relatively expensive. It also requires good designing and implementation arrangements.
Evaluation of I.12 based on data from soil monitoring programmes that are not specifically designed for the purpose of evaluation of RDP measures (e.g. regional soil monitoring programmes, research programmes, demo trials, etc.).	No (or very little) additional costs involved. Possibility to draw on (mainly) free-of-charge expertise and work. No need for undertaking any soil sampling and laboratory analysis. In most cases easy to arrange it from the institutional point of view (most of these programmes are financed by public money and are run by public research institutes or universities).	Available data will not necessarily fit I.12 requirements in terms of sampling depth and frequency, sampling density, etc. It could be quite challenging to properly 'transpose' it to RDP level. In most cases there will be no control group. The end result is likely to be methodologically questionable and not quite reliable.
Evaluation of I.12 based on qualitative analysis using Focus Groups, Delphi, MAPP, etc.	An elegant evaluation method in case no soil monitoring programme exists. Can be organised relatively quickly, easily and cheaply. Qualitative analysis stimulates interaction among stakeholders, enables information sharing and builds capacity among a wide group of people.	The evaluation will only be based on human observations, with no hard facts and figures based on laboratory soil analysis. The end result will be a qualitative guess/information, rather than an objective quantitative measurement.

Preconditions for applying Approach A and B

> **For the application of the suggested approaches, the MAs may consider the following issues:**
>
> - Is there a (sound) soil monitoring programme, which includes SOM analysis at RDP level? Are the monitoring data systematically collected/ easily available? (If so, the evaluators should be able to assess I.12 relatively easily).
> - If no sound soil monitoring programme is in place, what are the options for evaluators? E.g. they can be expected to spend more time and resources to assess I.12. In this case the following has to be checked:
>   - LUCAS data on SOM in arable fields – especially the density of sampling. Be aware that data down-scaling from LUCAS can be difficult both from a methodological and technical point of view.
>   - If other soil monitoring programmes (e.g. regional or national programmes are set-up to monitor erosion, nutrients content in soil, etc. – or demo trials/research programmes) have data on SOM, this data should be used, too.
>   - If no soil monitoring programme exits at all, evaluators should undertake qualitative analysis using Focus Groups, Delphi, MAPP, etc. They have to make sure to include a representative sample of stakeholders and make sure that these will take part in the analysis (sometimes it can be difficult to motivate them to take part in focus groups, etc.).
> - Is there the requirement for evaluators to demonstrate appropriate skills and methodology to assess I.12 as well as knowledge on existing information on SOM sampling frequency, sampling density, depth and laboratory methodology used?

4.5.4 Adequateness of suggested evaluation approaches

The above described evaluation approaches are in the following discussed as regards their adequateness in fulfilling the evaluation quality criteria: rigour, reliability, robustness, validity, transparency, credibility, practicability and cost effectiveness. The definitions of the quality criteria are provided in Table 2.

Table 11. **Adequateness of the proposed evaluation approaches for the assessment of CAP common impact indicator I.12 - Soil organic matters in arable land**

Quality Criteria	Approach A (optimal) Micro-level SOM assessment based on 0 – 60 cm soil depth	Approach B (alternative) Micro-level SOM assessment based on simplified soil monitoring programmes (0 – 20 cm) Macro-level SOM assessment based on LUCAS database
Rigour	**Approach A is rigorous**. It can produce reliable findings because it is based on a comprehensive set of indicators and soil analysis results obtained from the sufficient soil depth.	**Since** soil analysis results under this approach are obtained from the upper soil layer (0 – 20 cm) **it** will not be able to produce reliable findings in deep soils. Therefore, the **rigour of Approach B depends on the depth of the upper soil layer**
Reliability	**Approach A is reliable.** Especially if soil samples are taken more than once a year.	**Approach B is partly reliable.** The reliability also depends on the chosen pathway (see Figure 13 in PART II - logic model). Pathway 1 is more reliable than Pathway 3.

Quality Criteria	Approach A (optimal) Micro-level SOM assessment based on 0 – 60 cm soil depth	Approach B (alternative) Micro-level SOM assessment based on simplified soil monitoring programmes (0 – 20 cm) Macro-level SOM assessment based on LUCAS database
Robustness	**Approach A is robust,** when there is a sufficiently robust dataset both for beneficiaries and comparison group.	**Approach B is partly robust**. Its robustness depends on the chosen pathway (see Figure 13 in PART II - logic model). Pathway 1 is expected to be more robust than pathway 3, which in many Member States will be critical because of the low density of LUCAS soil samples.
Validity	**Approach A is valid.** The results should be quite accurate, comprising minor/usual errors arising from in-field and laboratory analysis.	**Approach B is partly valid.** Its validity depends on the chosen pathway (see Figure 13 in PART II - logic model). Pathway 1 is expected to be more valid than Pathway 3.
Transparency	**Approach A is transparent**. Soil sampling and laboratory testing methods applied, as well as used indicators being known, they can be checked, followed and replicated.	**Approach B is transparent**. The approach follows one of the three indicated (see Figure 13 in PART II - logic model). Soil sampling and laboratory testing methods applied, as well as indicators used in each pathway are known, can be checked, followed and replicated.
Credibility	**Approach A is highly credible**. Soil sampling should be made according to JRC recommendations and soil analysis following an ISO standard. The method can generate un-biased findings.	**Approach B is not credible**. None of the three pathways are credible because they are based on the soil depth of only 0-20 cm (see Figure 13 in PART II - logic model), containing only a minor portion of carbon in the soil.
Practicability	**Approach A is practical.** It can be applied without adverse consequences and will provide straight-forward results.	**Approach B is practical**. Its methods and estimation procedures are quite simple and easy to follow.
Cost-effectiveness	**Approach A is costly,** because it relies on a robust soil monitoring programme which can be quite resource consuming.	**Approach B is cost-effective.** It is less demanding and costly in terms of soil sampling, laboratory analysis and data analysis than for approach A.

4.5.5   Dos and don'ts

Dos

- Set up a soil monitoring programme, with several samplings per year- if possible, as long as the project runs (e.g. by using technical assistance).
- Follow best practices regarding soil sampling and laboratory analysis. Solely use national laboratories accredited by authorities for soil sampling and analysis.

Don'ts

- Forget to record soil bulk density for each sample of soil analysed, essential for calculating the SOC concentration in g/kg and for determining C stocks in soil, and their sequestration potential.
- Expect SOC to be noticeable before several years of implementation of RDP measures, conserving carbon.
- Expect LUCAS database to provide much data at NUTS 3 or lower level.

4.6 CAP common impact indicator I.13

The following sections of the Technical Annex are related to Chapter 2.8 in PART II of the Guidelines.

4.6.1 Additional indicators (examples)

To assess the RDP effect on soil erosion, the stakeholders may also use additional indicators. The European Soil Data Centre (ESDAC) provides a wide range of soil erosion indicators and relevant maps that are all based on simulations resulting from the initial 2009-2012 European-wide topsoil survey.[86] The indicators include wind erosion, and the RUSLE components, i.e. soil erodibility, the cover-management factor and the support practice factor that are briefly presented in Table 13. These additional indicators are important because they quantify wind erosion, currently covered by I.13, certain risk (erodibility) and support factors. The indicators show the areas already suffering from erosion or the areas at 'high risk', as well as the effect of management and support practices. Additional indicators are presented in Table 12.

---

[86] All simulations are based on the Revised Universal Soil Loss Equation (RUSLE) model (RUSLE 2015) with extreme precision (resolution up to 100m and Digital Elevation Model (DEM) at 25m). Information on the RUSLE can be obtained at: http://www.sciencedirect.com/science/article/pii/S1462901115300654 and at AEI21 fiche. Data are available upon request from: https://esdac.jrc.ec.europa.eu/content/soil-erosion-water-rusle2015

Table 12. Examples of additional indicators related to soil erosion by water

Indicator	Measurement Unit	Use of additional indicator	Data sources and frequency of collection
Wind erosion	Wind erosion is measured in Megagrams (Mg) per hectare per year (Mg ha-1 yr-1).	Wind erosion is an additional form of erosion which affects temperate climate areas of northern European Member States and semi-arid areas of European Mediterranean Member States. Wind erosion can have an average of well above 2 Mg ha^{-1}yr^{-1} with local maxima that are above 20 Mg ha^{-1} yr^{-1} (High Erosion).[87]	ESDAC, based on CORINE 2006 land cover, covers all EU territory, simulation published in 2017. Not known if it will be replicated with more recent data.
Soil Erodibility factor (K-factor)	K-factor is a pure number ranging from 0.02 (lowest erodibility) to 0.69 (highest). All other factors being equal, the higher the soil erodibility value, the greater the susceptibility of the soil to rill and sheet erosion by rainfall.	Soil Erodibility (also called the K-factor) is an indicator showing the intrinsic susceptibility of a soil to erosion by runoff and raindrop impact. Useful in planning and targeting 'high risk' areas. The mean K-factor for Europe was 0.032 in 2009.[88]	ESDAC, simulation based on 2009 LUCAS Soil, Bulgaria, Romania and Croatia are not included, simulation published in 2014. Different soil erodibility datasets are available through ESDAC's datasets. Not known if it will be replicated with more recent data.
Cover-management factor (C-factor)	The C-factor, is the ratio of soil loss from an area with specified cover and management to that from an identical area under the tilled continuous fallow Unit Plot conditions. The C-factor ranges from a value of zero, for completely non-erodible conditions, to 1.0 for the worst-case Unit Plot conditions. C. The mean C-factor in the EU is estimated to be 0.1043, with forests at 0.00116, and arable lands and sparsely vegetated areas at 0.233 and 0.2651, respectively.	The cover-management factor (C-factor) is the one that policy makers and farmers can most readily influence in order to help reduce soil loss rates. The C-factor quantifies the effects of management practices on soil erosion reduction.[89] Conservation practices reduce the C-factor by on average 19.1% on arable land.	ESDAC, simulation based on CORINE 2006 land cover and Eurostat agri-environmental indicators, covers all EU territory, simulation published in 2015 at the NUTS 2 level. Not known if it will be replicated with more recent data.
Support practice factor (P-factor)	P-factor is a pure number, part of the RUSLE. The P-factor is the ratio of soil loss with a support practice like contouring, strip cropping, or terracing to that with straight-row farming and down slope. The mean P-factor in the EU is estimated to be 0.9702.[90]	An indicator capturing the impact of support practices such as contouring, stone walls (also known as terrace sub-factor), and grass margins (also known as strip cropping sub-factor and buffer strips.[91] The support practices accounted for in the P-factor reduce the risk of soil erosion by 3%.	ESDAC, simulation based on 2012 LUCAS Soil, Croatia not included, simulation published in 2015. The P-factor dataset is available in raster format for the EU at 1km resolution or as a Mean P-factor at NUTS 2 level. Not known if it will be replicated with newer data.

---

[87] 'A New Assessment of Soil Loss Due to Wind Erosion in European Agricultural Soils Using a Quantitative Spatially Distributed Modelling Approach, Land Degradation and Development.', Vol. 28, pp. 335–344, Borrelli P. et al., (2016). Accessible at: http://onlinelibrary.wiley.com/doi/10.1002/ldr.2588/full. Data accessible upon request at: https://esdac.jrc.ec.europa.eu/content/Soil_erosion_by_wind

[88] 'Soil erodibility in Europe: A high-resolution dataset based on LUCAS, Science of the Total Environment.', Vol. 479–480, pp. 189–200, Panagos P. et al., (2014), available at: https://www.sciencedirect.com/science/article/pii/S0048969714001727. Data accessible after request at: https://esdac.jrc.ec.europa.eu/content/soil-erodibility-k-factor-high-resolution-dataset-europe

[89] 'Estimating the soil erosion cover-management factor at the European scale, Land Use Policy.', Vol. 48, pp. 38–50, Panagos P. et al., (2015), accessible at: https://www.sciencedirect.com/science/article/pii/S0264837715001611. Data accessible after request at: https://esdac.jrc.ec.europa.eu/content/cover-management-factor-c-factor-eu

[90] ESDAC for the K, C and P factors can be used to provide average values for countries and/or regions.

[91] 'Modelling the effect of support practices (P-factor) on the reduction of soil erosion by water at European scale, Environmental Science and Policy.', Vol. 51, pp. 23-34, Panagos P. et al., (2015), accessible at: https://www.sciencedirect.com/science/article/pii/S1462901115000611. Data accessible after request at: https://esdac.jrc.ec.europa.eu/content/support-practices-factor-p-factor-eu

4.6.2 Approach A

Micro-level assessment

Access to data and its quality and creation of a consistent database and data infrastructure

The evaluators may be able to use three distinct sources of data:

- Micro data of soil erosion by soil surveys maintained by the MA or the Member State that can support the setting-up of comparison groups;
- Micro data mentioned above, of soil erosion maintained by the LUCAS Soil database that can support the setting-up of comparison groups;
- Micro data that can be collected specifically for the evaluation.

If it is decided that data will be collected through a survey, the evaluator must be aware that a field survey cannot collect data that directly estimates the impact indicator, i.e. soil loss in tonnes per hectare and per year. Through a field survey one may be able to record either a visual inspection of erosion or some proxies showing if erosion processes are confronted. A way to address this issue is to assume that the erodibility factor (K-factor) is the same to all agricultural holdings in an area and that GAEGs (P-factor) apply to all of them equally. The only factor that can really make a difference on soil erosion is the C-factor (cover management). The C-factor can be estimated for each farm by simple arithmetic using field and monitoring data, following the simple methodology described in Panagos et al. (2015).[92]

Selection of a counterfactual option and micro-level method

In the case of the micro assessment, the control group is made up from farm holding that are not supported by the RDP. The micro-level assessment is conducted in the following distinct steps (Figure 15 in PART II Chapter 2.8.1):

Step 1: *Obtain the data for I.13-1 impact indicator from Eurostat* and estimate by GIS overlay methods indicator I.13-2. This step is depicted in the blue coloured boxes of above mentioned Figure 15.

Step 2: *Retrieve the data for result/target indicator R10/T12* and all available monitoring data, that will reveal the number of supported agricultural holding population and the variety of measures used within the RDP's intervention logic. Collect IACS/LPIS data to examine the spatial coverage of the supported agricultural holdings. Retrieve soil and land use maps to examine the degree of environmental heterogeneity in the areas covered by the RDP's measures. This step is depicted in the blue coloured boxes of above mentioned Figure 15.

Step 3: *Decide if the number of supported agricultural holdings (from Step 2 above) is sufficient* for carrying out a proper evaluation. Examine the possible heterogeneity of the RDP areas (from Steps 1 and 2 above).

Step 4: *Set up the counterfactuals*

a) Comparison groups will be created and the flow of operations from the top dark green box of above mentioned Figure 15 down to the pink boxes will be followed leading to Approach A.
d) Decide on the group or sub-groups of supported holdings (from Step 2 above) and of the control group, i.e. holdings which did not receive support from the RDP.
e) Decide on the spatial coverage of the survey as informed by Step 2 above. Decide on sample sizes depending on the method to be employed, the cost of sampling and the available budget.

Step 5: *Seek any alternative source* that can complement the existing sampling points from ESDAC, including national and local sources.

---

[92] 'Estimating the soil erosion cover-management factor at the European scale, Land Use Policy.', Vol. 48, pp. 38–50, Panagos P. et al., (2015), accessible at: https://www.sciencedirect.com/science/article/pii/S0264837715001611

Step 6: *Design a questionnaire* that will capture soil erosion through the C-factor.[93]The questionnaire should include questions to capture the slippage effect, i.e., the fact that conservation measures on part of the land may release production factors (especially labour) and intensify production on the remaining land. This is very similar to the leverage effect. Questions that capture any transverse effects and especially the effects of M01, M02.1 and M16.5 are also to be included.

Net impact assessment at micro-level

Step 7: *Analyse the data by applying an adequate statistics-based method.* We suggest:

a)   Use Propensity Score Matching which completely eliminates selection bias, if you have an adequate number of non-supported holdings (counterfactual) on which to match the supported holdings.

b)   Use simple regression on the C-factor with carefully chosen control variables that will reduce (not eliminate) selection bias. If you suspect selection bias to be strong, consider alternative formulations of the regression equation, such as a two-stage Heckman type model estimated with Full Information Maximum Likelihood (FIML). Simple regression approaches can be applied to small sample sizes. Two stage equation formulations with one participation equation and one impact equation are more demanding in terms of both the sample size and the statistical skills.

c)   Employ Instrumental Variables analysis that deals better with selectivity, but is more demanding econometrically and needs the existence of good instruments.

Step 8: *If the sample data cover has at least two distinct time periods*, i.e., if the evaluator can retrieve information for calculating the C-factor (from monitoring or other locally available data) before and during the programme's operation (or at the end), the methods in the previous step, and especially the matching algorithm, can be coupled with Difference in Differences (DiD) methodology for more robust results as concerns sample selection (Point 7 of above mentioned Figure 15).

Step 9: *Estimate the Average Treatment Effect on the Treated* (ATT) and compute the RDP's net direct effect coefficient on the C-factor.

Step 10: *In case of the K-factor, slope length and steepness as well as the rainfall erosivity and support practices (GAEGs) equally* affect RDP supported, and RDP non-supported farm holdings, the net direct coefficient estimated in the previous step is also the net direct effect coefficient of the soil erosion by water. For example, you may find that in your country or region through the C-factor, soil erosion is reduced by 25% on supported holdings and by 15% on matching non-supported holdings. The deadweight is 10%, which is still reasonable. Most probably, the RDP has focused on soil erosion prone areas. In these areas it is likely that all farmers undertake a minimum of protective management measures, e.g. all farmers maintain winter crop covers, and RDP supports additional efforts, e.g. reduced tillage.

Step 11: *If there are indications of important indirect effects* either on supported or non-supported agricultural holdings due to the application of soil conservation measures, which is rather unlikely, these should be treated separately. Qualitative methods and especially in-depth interviews with soil experts and agronomists will help the evaluator to highlight/identify such possible effects.

Step 12: *Aggregate the results and estimate the effects of the RDP at macro level.* At this stage the evaluator has two options. If the survey is representative of the RDP territory, the evaluator can scale up the micro results for both supported and non-supported farm holdings. In the case where the survey is not representative of the whole territory but only of specific targeted areas, the evaluator can work differently. One solution is to stratify the survey ex post and assign weights to the cases to make them representative of the entire population. This is rather cumbersome and data/time demanding. Alternatively, apply the net direct effects coefficients on the total RDP effects, as these are captured by

---

[93] Non- RDP supported holdings should be sampled from holdings that have the same K-factor characteristics, slope length and steepness as the supported holdings. K-factor variables (the textural factor, clay, silt and very fine sand fractions and organic matter content) should be included later in the econometric models, either as good control variables (regression approach) or as matching indicators (matching algorithms approach).

the relevant result Indicator, and only estimate the total RDP net direct effect. Compare the total soil loss as estimated by I.10-1 with the RDP's net direct effects. Does the RDP have an important contribution?

Step 13: *Verify the results obtained* by this process with qualitative data obtained by interviewing experts and by reviewing published case studies carried out in the RDP territory or in other RDPs, facing similar agricultural conditions.

Macro-level assessment

Access to data and its quality and creation of a consistent database and data infrastructure

For all Member States, data on soil erosion are available at the NUTS 3 level while data on the C-factor are available at the NUTS 2 level from ESDAC. Data are currently available for 2012, a point in time that may be assumed to have marked the start of the programme. Depending on the release of the 2018 LUCAS Soil survey, the existence of national or regional soil databases, access to spatially aggregate data must be evaluated on a case-by-case basis. Some Member States have their own extensive soil databases, while for other Member States the NUTS 3 data provided by ESDAC are adequate to employ a statistics-based method. In addition, it is important that the evaluator examines whether spatial data can be coupled with an environmental database addressing the same spatial level.

Selection of a counterfactual option and macro-level method

Depending on data availability (especially if there will be data during or after the programme's end), the number of spatial units (NUTS 3 or smaller) and associated environmental and monitoring data, the evaluator can consider several options for constructing the counterfactual and the method for estimating net effects. If there are spatial units that are not supported by the RDP and can be matched to spatial units that are supported by the RDP, then a Propensity Score Matching algorithm on spatial data can be applied. In this exercise matching will consider the physical (especially climatic, sloppiness and soil texture) and agricultural characteristics.

However, as this is rather rare, the counterfactual can also be constructed by dividing all spatial units into supported and non-supported following a threshold of support. This can be done with the use of the GPSM methodology. The GPSM methodology applied to a spatial-regional level is analytically presented in Michalek (2012).[94]

Net impact assessment at macro level

The quasi-experimental GPSM method nets out the RDP effects and avoids selectivity bias. To the best of our knowledge there are no general equilibrium models addressing soil erosion.[95]

Micro-macro consistency and validation

We propose that micro assessment focuses on the C-factor, while macro assessment focuses on soil erosion as estimated by ESDAC or national surveys. Thus, for the net impact assessment, it is important that the results of both micro and macro level assessment are consistent, i.e., results of these assessments show the same trend in relation to impact, even though the evaluator used different indicators or even different methods for the assessments. Triangulation of net impacts with information from qualitative sources, published case studies or academic work, should be performed.

---

[94] Accessible at JRC's website:
http://publications.jrc.ec.europa.eu/repository/bitstream/JRC72060/rdi%2025419%20(web)%20final.pdf
[95] Panagos et al (2018) used a CGE model to address the cost of soil erosion by modelling soil erosion as a loss in crop productivity. The work is accessible at: http://onlinelibrary.wiley.com/doi/10.1002/ldr.2879/full

Pros and cons of using Approach A

Important feature	Pros	Cons
Propensity Score Matching, is the best method for dealing with selectivity bias, but it requires a larger sample size as concerns non-RDP participants in order to ensure that the largest proportion of RDP participants can be matched to a non-participant.	Best method for addressing selection bias.	Larger sample sizes and intermediate econometric skills are required.
Simple regression methods do not demand high econometric skills but their ability to reduce selection bias and produce unbiased net effects estimates depends on careful design of the survey, and collection of 'good' control variables. If simple regression methods are extended to two stage estimations with participation and impact equations, then the easiness of application is reduced but the selection bias is better addressed.	Easiness of application.	Possibility of bias is not eliminated.
Instrumental Variables approach is econometrically more demanding than a simple regression but, if there are suitable instruments, it deals well with selection bias.	Selection bias is adequately dealt with.	Intermediate econometric skills are required.
Generalised Propensity Score Matching for spatial-regional data	It does not require existing units that did not receive programme support. Very effective tool for finding counterfactuals.	Validity of results depend on inclusion of important observable characteristics which explain differences between spatial units and especially climate, physical and soil characteristics.

Preconditions for applying Approach A

In case the MA decides to carry out a survey of farm holdings the possibility to combine this with a survey for other environmental indicators should be examined. (The open issues have been discussed in similar boxes for GHG and ammonia emissions in Chapter 2.3.1, and water abstraction and GNB in Chapter 2.6.1 of these Guidelines). Specific issues to be considered are:

- Will it be an RDP territory-wide survey or will it target specific RDP 'high soil erosion risk areas'?
- Will it address a wide range of measures or will it be restricted to only a few (2-3) measures?
- How to deal with the risk of double-counting if there are many (and overlapping measures)?

4.6.3 Approach B

Micro-level assessment

In the case where there is no time or resources to set up counterfactuals and follow approach A, the evaluator can adopt a simple approach in which supported agricultural holdings are compared with a population's average that consists of both supported and non-supported holdings with or without 'before' and 'after' estimates (Boxes 9, 10 and 11 in Figure 15 of Chapter 2.8.1 in PART II). The population's average serves as a control group. This approach assumes that in the absence of the RDP, soil erosion by water would be the same as for an average of supported and non-supported agricultural households. This is done at the expense of robustness, and especially at that of bias, introduced from sample selection. This method is preferred as a naïve method from the alternative that would compare supported holdings with an arbitrarily chosen small sample of non-RDP supported holdings. Especially if the number of supported holdings is relatively small in a larger area and thus their contribution in formulating the population's average is quantitatively smaller, then the bias introduced is probably less. However, the evaluator should be aware that this is a biased procedure usually leading to an overestimation of the RDP's net effects.

With this approach, the evaluator retrieves information from monitoring data and, if needed, complements these data with some additional information. As such, the evaluator will be able to calculate the average C-factor of supported holdings. This can be compared with the average C-factor of the NUTS 2 area in which supported holdings are located. Non-supported holdings do not have to be mandated or sampled. Furthermore, if from application data the evaluator can calculate a before and after change in the C-factor, then a naïve DiD approach can be employed. This procedure should be complemented by a sensitivity analysis of the outcomes and qualitative data collected from in-depth interviews with soil experts, agronomists, and possibly a focus group that will shed light on the magnitude of indirect effects such as displacement, substitution and multiplier effects.

Macro-level assessment

At the macro level, this approach can take the form of a quantitative naïve assessment between spatial units and a national average. However, this cannot be recommended as a sound and robust method.

Pros and cons of using Approach B

Important feature	Pros	Cons
The method allows for simple group average comparisons	Easiness of application because 'population' average data can be estimated using GIS methods on the ESDAC raster.	Net effects are only approximate. Selection bias is not addressed.

More pros and cons in relation to using MAPP can be found in Chapter 2.3.1 of PARTII of these Guidelines.

4.6.4 Adequateness of suggested evaluation approaches

The above described evaluation approaches are in the following discussed as regards their adequateness in fulfilling the evaluation quality criteria: rigour, reliability, robustness, validity, transparency, credibility, practicability and cost effectiveness. The definitions of the quality criteria are provided in Table 2.

Table 13. Adequateness of the proposed evaluation approaches for the assessment of CAP common impact indicator I.13 - Soil erosion by water

Quality criteria	Approach A (optimal) Micro-level **Statistic based evaluation techniques** Macro-level **GPSM spatial econometrics supported by DiD**	Approach B (alternative) Micro-level **Naïve baseline or dynamic group comparisons** Macro-level **Quantitative naïve assessment between special units and a national average**
Rigour	**Approach A is rigorous.** For micro-level, the surveys, if they exist and are of adequate size and quality, can produce robust results depending on the methodology used. For macro studies, if the sample size is adequate and/or if coupled with a spatial environmental database produces exact findings.	**Approach B is partly rigorous.** Naïve comparisons between supported and average population can only show one trend and can produce a biased estimate. However, bias may not be that serious depending on the number of supported holdings out of total holdings in the region.
Reliability	**Approach A is reliable**, if the methods measuring soil erosion at micro-level remain the same over time. Currently soil erosion is the outcome of a simulation model. In the LUCAS Soil 2018 survey, soil erosion will be included as a site characteristic with visual inspection/measurement. Macro estimates will depend on harmonisation of simulated data resultant from the 2009-2012, 2015 and 2018 surveys.	**Approach B is partly reliable**. The reliability depends on the size of the selection bias that is introduced. If the areas selected for support by the RDP are areas under high erosion risk then, selection bias will inflate differences between groups or between groups and the national average.
Robustness	**Approach A is robust.** In case of micro estimates the robustness depends on the choice and availability of variables to control the matching algorithm or the regression. For macro estimates, the robustness is ensured if the number of spatial units is adequate.	**Approach B is partly robust.** The robustness depends on the size of the population of supported holdings.
Validity	**Approach A is valid.** Statistical estimates of errors are produced for micro studies and should always be taken into account when	**Approach B is partly** valid but logical and sound under certain circumstances. Depending on the number of cases,

73

Quality criteria	Approach A (optimal)	Approach B (alternative)
	Micro-level **Statistic based evaluation techniques** Macro-level **GPSM spatial econometrics supported by DiD**	Micro-level **Naïve baseline or dynamic group comparisons** Macro-level **Quantitative naïve assessment between special units and a national average**
	results are the subject of a sensitivity analysis. For macro-level the simulated results should always be validated by field visits.	statistical estimates of errors may not be possible to be estimated or their confidence intervals may be very broad.
Transparency	**Approach is transparent.** All data and methods are publicly available, and all processes are recorded and unified.	**Approach B is transparent.** All data and methods are publicly available, and all processes are recorded and unified.
Credibility	**Approach is credible.** But it must be taken into account that soil erosion processes are slow unless abrupt changes are taking place. For macro-level the selection bias is well handled by the proposed econometric methods. Superior information can point to areas where selection bias may be an issue	**Approach B is partly credible**, because selection bias will be present. The severity of bias depends on several factors and especially on the number of supported holdings.
Practicability	**Approach is practical.** Used methods and estimation procedures are fully practicable because they do not measure soil erosion as loss of soil which is very difficult. The method estimates the K-factor which is easily calculated from land cover information.	**Approach B is practical.** Used methods and estimation procedures are fully practicable and can be implemented without needing special skills or training.
Cost-effectiveness	**Approach can be costly.** This depends on the cost per sampled unit and the sample size. The ability to pre-process data for RDP supported holdings, and especially data for land use and cover crops, will reduce the cost.	**Approach B is cost effective.** It can provide a cost-effective way of assessing RDP impacts on restricting soil erosion under limited data availability at the expense of reliability and robustness.

4.6.5 Soil Erosion Data availability

LUCAS Soil is the database of the topsoil surveys carried out across the European Union to derive policy-relevant statistics on the effect of land management on soils, and it complements the Land Use/Cover Area frame Survey (LUCAS) point survey.

LUCAS Soil surveys have been carried out during the period 2009-2012 and 2015 at approximately 45,000 sites in Europe (22,000 in the 2009-2015 period and the rest in 2015). The 2009-2012 data, at sampling point level, are available upon request from ESDAC and the 2015 data are expected to be released in 2018. In the 2009-2012 and 2015 surveys, LUCAS Soil targeted physicochemical properties, including pH, organic carbon, nutrient concentrations and cation exchange capacity. Based on raw soil data from the 2009-2012 survey in combination with the Land Use/Cover Area frame Survey (LUCAS) and other available surveys, the Institute for Environment and Sustainability of the Joint Research Centre (JRC-IES) of the European Commission has carried out a number of simulations based on the RUSLE to estimate the K-, C- and P- factors which are available in raster format. These simulations will probably be repeated for the 2015 survey data to further refine the simulation estimates.

As from 2018, additional properties, including bulk density, soil biodiversity, specific measurements for organic-rich soil and soil erosion will be measured.[96] Assessment of soil erosion by water and wind will

---

[96] 'LUCAS Soil, the largest expandable soil dataset for Europe: a review, European Journal of Soil Science.', pp.1-14, Orgiazzi A. et al., (2017), accessible at: http://onlinelibrary.wiley.com/doi/10.1111/ejss.12499/full

be visual, and will collect information that will be used to improve soil erosion modelling and improve assessments of land degradation. Visual evaluation of the erosion will be carried out at approximately 26,000 sampling points and will indicate the type of erosion (i.e. sheet, rill, gully, mass movement, re-deposition and wind erosion), distance and direction from the LUCAS point, together with an estimate of the number of rills or gullies observed. This will allow an excellent simulation of soil erosion by water and wind and will allow an even more precise evaluation of erosion effects (spatially and numerically). However, the time when the 2018 survey point data will be released is not known yet. We urge evaluators to use this database and the results of these simulation models because data are harmonised across all Member States as concerns both the field collection (observation) procedures and their laboratory analysis and treatment as well as their simulation procedures based on the same supporting material (e.g., the LUCAS cover survey, the same DEM, etc.).

The LUCAS Soil data at sampling point level may, in certain circumstances, be enough for an RDP to set up comparison groups. In other words, the soil sample size in an RDP may be adequate and the spatial disposition of the sampling points may include a number of supported and non-supported farm holdings that allow the set-up of comparison groups. In that case, the evaluator may use the existing LUCAS Soil database and complete it with own observations on soil erosion along the lines suggested by the 2018 LUCAS Soil survey.

Apart from the LUCAS Soil database, some RDPs[97] or Member States maintain their own soil surveys or detailed soil maps that can assist and support the evaluation process.[98]

4.6.6 The Evaluation of the C-factor for arable land[99]

The evaluator can estimate the C-factor which captures the effects of land cover and of management practices (tillage, cover crops and residue management) for the areas targeted by the RDP with soil management/erosion contracts by following the procedure cited in Panagos et al (2015). The estimation of the C-factor shows the gross effects of the RDP for these three management practices. C-factor for arable land is estimated as:

$$C_{arable} = C_{crop} \times C_{management}$$

The crop factor (C_{crop}) is estimated as the weighted average of 17 different crops that may be present on a farm. The evaluator can adopt the factors per crop type from Table 1 of Panagos et al (2015). The management factor ($C_{management}$) quantifies the effect of management practices as:

$$C_{management} = C_{tillage} \times C_{residues} \times C_{cover}$$

The tillage coefficient of the 'management factor' of the C-factor equation is estimated as:

$$C_{tillage} = F_{conventional} \times 1.0 + F_{conservation} \times 0.35 + F_{No\ tillage} \times 0.25$$

with $F_{conventional}$ being the fraction of arable land with conventional tillage, and $F_{conservation}$ and $F_{No\ tillage}$ are the fractions of arable land with conservation tillage and No tillage at all. The residues coefficient of the 'management factor' of the C-factor equation is estimated as:

$$C_{residues} = 1 \times (0.88 \times F_{residues}) + (1 - F_{residues})$$

where $F_{residues}$ is the fraction of arable land treated with plant residues. The cover crops coefficient is estimated as:

---

[97] The Soil Information System for Lower Saxony (NIBIS) is a notable example providing an area database of soil maps, a borehole database, a laboratory database and a methods database. Even in this detailed database, however, soil water and wind erosion data are sparse and fragmented, and efforts have been taken to enrich information.
[98] ESDAC offers a very detailed section that provides links to websites from where soil related datasets at regional level could potentially be accessed and/or downloaded. This is accessible at: https://esdac.jrc.ec.europa.eu/content/regional-data
[99] This section draws exclusively on Panagos et al., (2015) at: https://www.sciencedirect.com/science/article/pii/S0264837715001611

$C_{cover} = 1 \times (0.80 \times F_{crop\text{-}cover}) + (1 - F_{crop\text{-}cover})$

where $F_{crop\text{-}cover}$ is the fraction of arable land to which cover crops are applied during winter or spring.

The coefficients suggested by Panagos et al (2015) are based on thorough literature search. However, the evaluator may have access to own or local estimates which may differentiate these coefficients to better reflect local conditions.
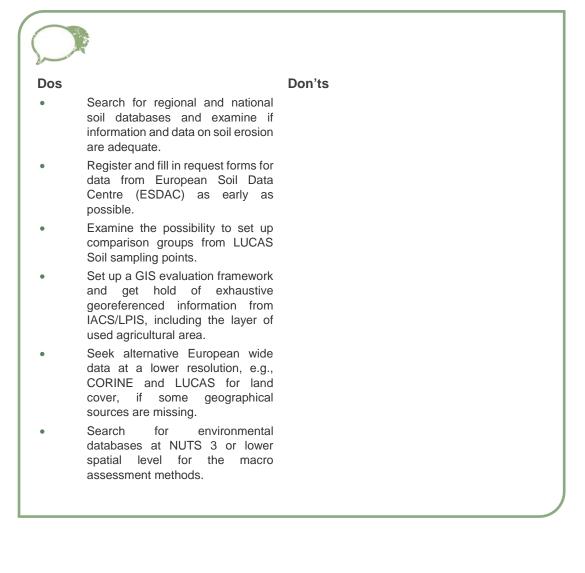
4.6.7 Control variables

The simple regression approach can be used for causal interpretation if the Conditional Independence Assumption (CIA) is valid. The CIA (also called selection on observables), is a core assumption implying that 'comparisons of average soil erosion reduction across supported (treatment) and non-supported (control) groups has a causal interpretation and selection bias disappears' conditional on observed characteristics. The choice of control variables is of crucial importance for the success of this method. The choice of the same variables is also crucial for the success of the matching algorithms (propensity score, etc.). A good and detailed knowledge of the process that determines participation or non-participation will assure that CIA holds and will refrain from the use of bad controls. Good controls variables are those that can be thought of as fixed at the time the support-no-support variable was determined. The holding's K-factor characteristics are typical good control variables and include the textural factor, clay, silt and very fine sand fractions and organic matter content, slope length and steepness. Most of them can be retrieved from soil maps/databases and DEMs. From the human capital characteristics, education is a good control variable, but training is not, especially when support is associated with obligatory training. Bad controls are variables that are themselves outcomes of the treatment variable. For example, irrigated area may be reduced because irrigation may accelerate erosion on high slope parcels. As a result, variables capturing the extent of irrigation should not be used as control variables. Failing to control on observables will introduce selection bias which, most often, will exaggerate the benefits of support.

Technical Literature

- Borrelli P. et al. (2016). A New Assessment of Soil Loss Due to Wind Erosion in European Agricultural Soils Using a Quantitative Spatially Distributed Modelling Approach, Land Degradation and Development, Vol. 28, pp. 335–344 Accessible at: http://onlinelibrary.wiley.com/doi/10.1002/ldr.2588/full
- Panagos P. et al. (2014). Soil erodibility in Europe: A high-resolution dataset based on LUCAS, Science of the Total Environment, Vol. 479–480, pp. 189–200, at: https://www.sciencedirect.com/science/article/pii/S0048969714001727
- Panagos P. et al. (2015). Estimating the soil erosion cover-management factor at the European scale, Land Use Policy, Vol. 48, pp. 38–50. Accessible at: https://www.sciencedirect.com/science/article/pii/S0264837715001611
- Panagos P. et al (2015). Modelling the effect of support practices (P-factor) on the reduction of soil erosion by water at European scale, Environmental Science and Policy, Vol. 51, pp. 23-34. Accessible at: https://www.sciencedirect.com/science/article/pii/S1462901115000611
- Orgiazzi A. et al. (2017). LUCAS Soil, the largest expandable soil dataset for Europe: a review, European Journal of Soil Science, pp.1-14. Accessible at: http://onlinelibrary.wiley.com/doi/10.1111/ejss.12499/full
- Panagos, P. et al. (2018). Cost of agricultural productivity loss due to soil erosion in the European Union: From direct cost evaluation approaches to the use of macroeconomic models. Land Degradation and Development, 1–14. http://onlinelibrary.wiley.com/doi/10.1002/ldr.2879/full

All suggested literature is in open access articles under the terms of the Creative Commons Attribution License and thus access to the content of the paper should not present any problem. Most of the papers also are accessible through ESDAC's web site.

4.6.8   Dos and don'ts

Dos

- Search for regional and national soil databases and examine if information and data on soil erosion are adequate.

- Register and fill in request forms for data from European Soil Data Centre (ESDAC) as early as possible.

- Examine the possibility to set up comparison groups from LUCAS Soil sampling points.

- Set up a GIS evaluation framework and get hold of exhaustive georeferenced information from IACS/LPIS, including the layer of used agricultural area.

- Seek alternative European wide data at a lower resolution, e.g., CORINE and LUCAS for land cover, if some geographical sources are missing.

- Search for environmental databases at NUTS 3 or lower spatial level for the macro assessment methods.

Don'ts

4.7 CAP common impact indicators I.14, I.15, I.16

The following sections of the Technical Annex are related to Chapter 2.9 in PART II of the Guidelines.

4.7.1 Approach A1 and A2

Both evaluation approaches are only carried out at macro-level, where the unit of analysis is rural areas within the RDP territory taking into consideration either NUTS 3 or LAU 2 level.

Macro-level assessment for Approach A1

Approach A is conducted in several steps which are described below:

Step 1: *Construct the model with the appropriate data:* In the case of the Recursive-Dynamic CGE the most demanding data needs are associated with model construction. The basis for a regional/rural CGE model is a mechanically-constructed SAM.[100] As information necessary for the SAM construction requires regional employment and accounting data at the sectoral level, it is advised that the CGE model is built at the level of RDP-specific rural NUTS 3 regions as defined by the Eurostat Urban-Rural typology.

- First, sectoral employment data is needed to downscale an available national Input-Output table (which should ideally be for a year close to the commencement of the RDP) for the programme-area level;
- Second, data needs for filling the inter-institutional and factor-institutions flows can (usually) obtained from regional accounts, household income and expenditure surveys;
- Third, the latter should also be used in the (very frequent) case where households are disaggregated to different types according to (e.g.) income levels.

Step 2: *Calibrate the model:* The calibration of the dynamic CGE model requires the specification of a wide range of production, trade and household consumption elasticities. When the analysis is at the regional level, often such elasticities are based on reviews of the relevant literature. The same holds for the definition of exogenous parameters which are often available at national level and hence, significant fine-tuning is often needed to downscale them. Last but not least, deep knowledge (and often expert opinion) is used in order to specify study-area specific closure rules on factor markets, government budget, the regional current account, and the investment and savings account.

In the case of measure-specific financial flows, the needed information is annual expenditure (which has to be converted into model baseline prices), as well as data on the sectoral targeting of flows for each measure.

Step 3: *Control model dynamics with appropriate adjustments:* To control model dynamics, a number of exogenous 'between period' adjustments on variables such as productivity growth, government spending, population and labour supply can be imposed in the Recursive-Dynamic version of the model. These adjustments should be imposed through the use of real data for the period 2014-2018 (in case the approach is applied for AIR 2019) and projections for the period for which real observations do not exist (e.g. in case of the RDP ex-post evaluation). Capital adjustment for each sector between periods is typically endogenous, with investment by commodity in the solution of the model in period *t-1,* used to update capital stocks before the model solution in period *t.* Assuming that the commodity composition of capital stock is identical across activities, the allocation of new capital across activities uses a partial adjustment mechanism, with those activities where returns are higher than average obtaining a higher than average share of the available capital.

Step 4: *Estimate the impact indicators with appropriate additional data:* CGE model outputs include RDP measure-specific annual impacts on employment, household income and GDP. Therefore, in

---

[100] For the popular GRIT regionalisation technique, see 'Regional Economic Planning: Generation of input-output analysis.', London, Jensen et al., (1979).

order to estimate the aforementioned impact indicators, the following additional data should be obtained:

- I.14: Study-area-specific changes in population aged 15-20 and over, since the start of the programming period, net of non-RDP measure effects. These estimates can be generated through the application of a qualitative method (see Section 2.9.4 of this Chapter);
- I.15 and I.16: Study-area-specific changes in total population, since the start of the programming period, net of non-RDP measure effects. These estimates can be generated through the application of a qualitative method (see Chapter 2.9.4);
- I.16: PPS conversion rates which are available from Eurostat.

To sum up, with the exception of RDP flows data, which should normally be available from programme-implementing authorities, data availability for the construction and calibration of the model is a rather case-specific issue. Research experience has shown that data availability varies considerably amongst case studies, leading (in the case of restrictions) to second-best choices in terms of designing a model with a more aggregate sectoral structure.

Pros and cons of using Approach A1

Important feature	Pros	Cons
Recursive-Dynamic CGE models are a rather advanced method for assessing development policy economic impacts, able to produce net effects estimates, due to their underlying behavioural characteristics and ability to generate counterfactuals.	Recursive-Dynamic CGE models can assess RDP impacts associated with increased capacity, in a robust manner, as they take into account the fact that most RDP projects/measures take more than one year to be completed. Further, estimates produced are net, as the CGE model impact estimates account for displacement, deadweight, secondary, unintended, multiplier and allocative efficiency effects of policy.	The method requires advanced technical skills. Data requirements can also be demanding and technical skills are required to specify data requirements in a cost-effective manner.

Preconditions for applying Approach A1

> The necessary preconditions can be understood by consulting Thurlow (2008) and Psaltopoulos et al. (2012) on the technical issues of the CGE model. There are three crucial issues to be considered:
>
> - Does the evaluator have enough human resources with experience on CGE modelling?
> - Is the close cooperation between the evaluator and the MA on the structure of the model (disaggregation of sectors, production factors, households) ensured? (Within this context, it is considered worthy to disaggregate agriculture into sub-sectors using FADN data for the baseline year of the SAM. Also, ideally, sectors targeted by the RDP should be explicitly included in the model).
> - Are the data for the baseline year for the SAM (i.e. CGE database) available? (the baseline year should correspond to the year before the start of the programming period - 2013. If there is no SAM available (from past efforts), the SAM model construction should be based on the available national IO tables for this year. If the analysis corresponds to the regional level (regional RDP), then the simple GRIT method should be applied in order to regionalise the national IO table.)

Macro-level assessment for Approach A2

Access to data and its quality & creation of a consistent database and data infrastructure

PSM and GPS methods are rather data demanding. Collected economic data should include all relevant information on programme-supported regions and control regions covering periods 'before' and 'after' the implementation of the programme. This data must include their economic and structural characteristics and their performance (including data on impact indicators to be assessed, i.e. I.14, I.15 and I.16) '"'The most important challenge here is to collect in order to construct meaningful control groups. In this regional / macro-economic analysis, counterfactual design requires data to be collected for individual rural regions within RDP (see above). In most cases, detailed (secondary) data can be obtained from respective statistical offices. However, data on small communities or villages may be collected through surveys.[101]

For the application of binary PSM, the availability of this data at rural LAU2 level is an important condition for the specification of control regions and hence, for the application of the method. Other data needs to include the distribution of measure-specific data on the allocation of funds at the same geographical level.

Selection of a counterfactual option

Approach B uses a quasi-experimental technique based on counterfactual analysis involving comparison of rural territories.[102] The workflow involves the following steps.[103]

Step 1: *Define model choice, programme-supported and control regions:* As already mentioned, in the case of the application of PSM, programme-supported regions and control regions should ideally be LAU 2 rural areas as defined by the Eurostat urban-rural typology. If there is no data available at LAU 2, the use of the NUTS 3 level Eurostat specification of rural areas is suggested. In case of the latter, Generalised PSM (GPS) should be applied.

---

[101] In some countries the micro-spatial grid data exists (e.g. one sq. kilometre resolution), such as microm in Germany and Austria. If evaluators have access to those data, they can offer considerable insight, because of the high resolution compared to rough aggregate averages at municipality or regional level.
[102] Lokshin and Yemtsov (2005); Michalek (2008).
[103] For a detailed description, see Section 3.3.2 of Metis/WIFO/AEIDL (2014).

First, programme-supported and control regions are specified on the basis of RDP measure-specific support outlays. In the application of a binary PSM to the estimation of RDP impacts,[104] control regions are defined either by their non-participation in the measure(s), or by setting an arbitrary low level of measure-specific support received (i.e. intensity of programme/measure exposure). As the measurement of programme/measure intensity per region may be problematic for small regions, alternative participation methods such as programme/measure exposure per capita or per square km can be alternatively applied.

Step 2: *Variable choice:* Second, the impact indicator targeted by the analysis is defined (in this case, I.14, I.15, I.16) and the impact of RDP measure(s) implemented in specific rural regions is analysed in both programme-supported regions and control regions, prior to the programme and after it, by applying a combination of PSM and Difference-in-Differences methods.

Step 3: *Conducting the assessment:* Third, the evaluation of RDP impacts is performed on the basis of estimated Average Treatment Effects (effect on an average region randomly selected from the pool of measure participants and non-participants), Average Treatment on Treated (effects on regions which participated in the measure) and Average Treatment on Untreated (effect of the measure on regions that did not participate).

Step 4: *Sensitivity analysis:* As a fourth step, sensitivity analysis (e.g. rosenbaum bounding approach) is carried out, to assess the possible effects of un-observables on obtained results.

Step 5: *Estimation of overall impact (in case of application of GPS):* Finally, in the case of GPS, the overall impact of the measure support is estimated through a dose-response function and derivative dose-response functions.

Pros and cons of using Approach A2

Important feature	Pros	Cons
Propensity Score Matching is an advanced and effective tool applied in development programme evaluation, which enables the construction of counterfactuals and produces net policy effects at the programme area level.	PSM incorporates numerous general equilibrium effects of a programme (e.g. substitution, multiplier, etc). Estimated net effects computed are specific to a wide range and can be positive or negative, primary and secondary, expected or unexpected, intended or unintended.	PSM requires abundant and good quality data (at regional level) and considerable technical skills.

Preconditions for applying Approach A2

The following shall be considered if approach A2 is requested:

- Are there data (on impact indicators to be assessed and other socio-economic characteristics) at LAU2 level available? (If this data is available PSM can be applied. If this data is available only at NUTS 3 level, then the ToR should be specific to the application of the Generalised PSM Method. Various sources as described in Annex 4.7.1 of these Guidelines can be checked).
- Are there sufficient human resources with experience on econometric analysis?

---

[168] Metis/WIFO/AEIDL (2014).

4.7.2 Approach B

Macro-level assessment

As for Approach A, the alternative evaluation approach (Approach B), is only carried out at the macro-level, where the unit of analysis should be defined at the NUTS 3 level for rural areas within the RDP territory. The steps of IO analysis are presented below:

Step 1: *Construct the model with appropriate data:* In the case of the IO model, model construction data needs are not excessive. They include a national IO table close to the start of the programming period, and sectoral employment data, the latter being used to downscale the IO table at rural case-study level, through the application of regionalisation techniques such as GRIT. Programme measure-specific data required, is annual expenditure data disaggregated by type of expenditure (e.g. construction, machinery, etc.) which is necessary to allow the specification of sectoral shocks on final demand. In turn, data on measure-specific adjustment of productive capacity is needed (e.g. data on changes in GVA or employment), in order to carry out the capacity-adjustment analysis part. Normally, this programme data should be available from RDP implementing authorities.

Step 2: *Select a counterfactual option and micro-level method:* However, as already noted, the shortcoming of the IO method to directly deal with counterfactuals means that a separate counterfactual analysis should be carried out in order to estimate net capacity-adjustment effects. Hence, econometric counterfactual analysis at micro level should provide estimates on measure-specific changes in GVA or employment.[105] In the case of socio-economic impact indicators, it is suggested that counterfactual analysis can be rather easily implemented for private investment measures specific to the three FA of Priority 6. However, it is judged that in order to capture the 'soft' perspective of M01, M02, M16, M13 and M19.1, M19.3 and M19.4 and the rather wider (and hard to capture) impacts of M07, a qualitative analysis is needed in order to approach the counterfactual.

Step 3: *Estimate the policy impacts:* IO modelling incorporates sectoral analysis into a macroeconomic framework, thus creating a basis for an evaluation of sectoral and/or investment policies for national or regional goals such as GDP and employment. An IO model can be used to estimate the indirect and induced effects of a change in the final level of demand for the output of a particular sector (impact analysis). These effects may be measured as output, income, and employment changes, calculated using sectoral multiplier coefficients, which express the ratio of total effect to the initial change in demand. Impact information is available in disaggregated as well as total form, and policy makers can thus be provided with information on which industries or sectors are impacted by a specific shock.

The transformation of an IO table into an economic model facilitates the analysis of economy-wide impacts of exogenous demand shocks, including development policy interventions. Two types of effect can be modelled.

- First, investment effects can be estimated, financial flows associated with specific RDP measures can be inserted into the IO model in the form of sector-specific exogenous demand shocks. Subsequently, following the traditional Leontief procedure, economy-wide growth generating impacts are estimated for each RDP measure.
- Second, capacity-adjustment effects, which are economy-wide effects of RDP projects at the operation stage can be estimated through the application of the 'mixed exogenous/endogenous variable version of the Leontief Model.'[106]

Last, but not least, similar to the case of the CGE models, a counterfactual 'exercise' is necessary in order to net-out changes in population levels, and a qualitative analysis is proposed to this end, with the use of focus groups or Delphi.

---

[105] See Metis/WIFO/AEIDL (2014).
[106] For details, see Psaltopoulos et al. (2004); Metis/WIFO/AEIDL (2014).

Step 4: *Estimate the impact indicators with appropriate additional data:* Finally, similarly to the CGE models, IO model outputs include RDP measure-specific impacts on employment and income. Hence, additional data is required in order to estimate indicators I.14, I.15 and I.16. This includes study-area-specific changes in population aged 15-20 and over (I.14), and in total population (I.15 and I.16) since the start of the programming period, net of non-RDP measure effects; and PPS conversion rates (I.16).

Pro and cons of using Approach B

Important feature	Pros	Cons
Input-Output Analysis is a general equilibrium quantitative technique, widely applied to the estimation of policy impacts. It can be used to estimate measure/project effects associated with both investment activity and capacity adjustment.	Conditional to the provision of counterfactual data on measure-specific net adjustment of productive capacity, IO analysis can estimate net impacts of RDP measures.	The restrictive assumption of the IO method (e.g. fixed input structure, unlimited capacity of primary factors, no price effects) result into the overestimation of policy impacts.
	Intermediate technical skills are required to apply the method.	If IO is not combined with counterfactual data, it leads to the generation of naive estimates.

Preconditions for applying Approach B

The preconditions can be understood by consulting the final report 'Investment support under Rural Development Policy, Metis/WIFO/AEIDL (2014)'. There are three crucial issues to attend to:

- Does the evaluator have sufficient human resources with experience on quantitative analysis?
- Is the close cooperation between the evaluator and the MA ensured in the structure of the model (disaggregation of sectors)? (Within this context, it is considered worthy to disaggregate agriculture into sub-sectors through using FADN data for the baseline year of the IO model. Also, ideally, sectors targeted by the RDP should be explicitly included in the model).
- Are the data for the baseline year for the IO model available? (The baseline year of the IO model should correspond to the year before the start of the programming period - 2013. If the analysis corresponds to the regional level (regional RDP), then the simple GRIT method should be applied in order to regionalise the national IO table).

4.7.3 Qualitative methods for counterfactuals

The use of qualitative analysis is needed to complete the assessment of CAP impact indicators I.14, I.15 and I.16 as provided with previous quantitative methods. Two methods are proposed: the MAPP and the Delphi methods.

The Delphi method [107] is a well-known, traditional expert panel method that can be used to assess inter alia the probability and intensity of effects of different interventions and measures.

MAPP is an innovative focus group method for the assessment of impacts of programmes and projects that has only been used in recent years in rural development evaluations. [108] It is best applied at local

---

[107] Delphi method
[108] Ex-post evaluation of RDPs 2000-2006 and Study on Investment Support under rural development policy (Metis, 2014), where a detailed description can be found.

and regional level, but since the indicators to be estimated here could also correspond to the national/rural level, it is proposed to implement the method in several rural regions and then extrapolate.

MAPP will be used with two purposes:

- To assess net population changes for the CGE and IO models; in this case the qualitative methods can be used to estimate study area specific changes in total population and by age group, since the start of the programming period, net of non-RDP effects.
- To triangulate the evaluation findings obtained with above the quantitative methods; the recommended steps in using the MAPP method for assessment of socio-economic indicators are as follows:

Step 1: *Select the regions* as suggested for the quantitative methods. Ideally, the whole RDP territory should be covered. If it is not possible to conduct several focus groups, then a limited number of regions are selected, based on their population size, so that they can be representative of the RDP territory.

Step 2: *Select the RDP measures* as they are depicted in the intervention logic of Figure 16 in PART II, Chapter 2.9.1, notably M01, M02, M04, M06, M07, M08, M16 and M19.

Step 3: *Select the indicators to be assessed* with the MAPP. They should be:

- population, 15-20 and over as well as total population (for the CGE and IO models);
- GVA, employment and poverty rate for capturing the soft or wider effects of the relevant measures.

Step 4: *Select the participants,* e.g. representatives of beneficiaries and non-beneficiaries should be invited to the focus groups.

Step 5: *Select the MAPP tools* from the range of MAPP tools, the relevant ones here are:

- the trend analysis tool, where detailed development trends are evaluated over the same time period according to a number of pre-defined indicators (the ones selected in Step 3);
- the influence matrix, which helps evaluate the influence of all interventions (RDP measures as well as other interventions in the area, to net out the RDP effects) on each indicator;
- the impact profile, which summarises the scale of the impact on each indicator from different measures/interventions and explains the main influences.

Step 6: *Report on MAPP results,* which are twofold:

- an estimated Figure for the net population change over the programme/evaluated period;
- validated findings from the quantitative assessment.

4.7.4 Adequateness of suggested evaluation approaches

The above described evaluation approaches are in the following discussed as regards their adequateness in fulfilling the evaluation quality criteria: rigour, reliability, robustness, validity, transparency, credibility, practicability and cost effectiveness. The definitions of the quality criteria are provided in Table 2.

Table 14. **Adequateness of the proposed evaluation approaches for the assessment of CAP common impact indicators I.14 – rural employment rate, I.15 – degree of rural poverty and I.16 – GDP per capita**

Quality criteria	Approach A1 (optimal)	Approach A2 (optimal)	Approach B (alternative)
	Macro-level **Recursive-Dynamic CGE Models**	Macro-level **Propensity Score Matching (PSM) and GPSM**	Macro-level **Input-output (IO) Analysis**
Rigour	**Approach A1 is rigorous** as it is based on causal analysis linked to well-developed economic theory. However, further elaboration is needed to produce values	**Approach A2 is rigorous.** It produces exact findings because it is based on well-developed statistical theories about causal	**This approach is partly rigorous** due to its reliance on linear relationships determining economic behaviour. Also, further elaboration is needed to

Quality criteria	Approach A1 (optimal)	Approach A2 (optimal)	Approach B (alternative)
	Macro-level **Recursive-Dynamic CGE Models**	Macro-level **Propensity Score Matching (PSM) and GPSM**	Macro-level **Input-output (IO) Analysis**
	on I.14, I.15 and I.16, which involves the estimation of RDP-area changes in population levels attributed to the RDP.	effects and concepts on how to measure such effects.	produce values on I.14, I.15 and I.16, which involves the estimation of RDP-area changes in population levels attributed to the RDP.
Reliability	**Approach A1 is reliable** and can produce stable and consistent results if the constructed is robust. Model robustness can be checked through sensitivity analysis.	**Approach A2 is reliable.** If data are available, it is possible to repeat all analytical steps and produce similar results during repeated observations in identical conditions.	**This approach is reliable** to a certain extent, due to the inherent tendency of the IO methodology to overestimate policy impacts.
Robustness	**Approach A1 is robust,** if the model is correctly parametrised and if this is tested through sensitivity analysis.	**Approach A2 is partly robust.** Robustness of results shall be checked by applying sensitivity analyses, e.g. determining the influence of an 'unobservable' on obtained results.	**This approach is robust** and results produced are stable and resilient to small changes in policy shocks.
Validity	**Approach A1 is valid.** It is able to capture a wide range of economic effects.	**Approach A2 is partly valid**, as the method is not driven by principles of economic behaviour.	**This approach is valid** in comparative terms. Errors can appear on estimates if judged on absolute terms.
Transparency	**Approach A1 is not very transparent,** as both the model database construction and parameterisation processes can be very complex.	**Approach A2 is rather transparent**, as the code of the analysis is written in a given programming language, users know exactly its main elements, structure, parameters, rules and functional responses.	**This approach is very transparent,** as it is built on simple elements, structure, parameters, rules and functional responses.
Credibility	**Approach A1 is partly credible,** as it cannot perfectly isolate the impact of programme effects from those associated with economic structures and inter-relationships. To enhance credibility, sensitivity analysis specific to model shocks must be carried out.	**Approach A2 is credible.** It allows to isolate programme effects from other factors and eliminate the selection bias.	**This approach is credible**, if it is combined (fed by) a rigorous counterfactual analysis at micro level.
Practicability	**Approach A1 is partly practical.** The method can be used in case of good data availability.	**Approach A2 is partly practical.** The method is demanding in terms of data, time and depends on the evaluator's skills.	**This approach is practical** and can provide a quick fix for estimating impacts for the 2019 AIR.
Cost-effectiveness	**Approach A1 is cost-effective** only if the modeller is based on data availability in order to determine model structure.	**Approach A2 is cost-effective** if the necessary data is available.	**This approach is very cost-effective,** since it is not demanding in terms of the amount of data needed for the assessment of policy effects.

4.7.5   Dos and don'ts

Dos

- Map available data and subsequently decide the model structure.
- Draw a line (cost/benefit) of searching for and using model construction data.
- Make sure that data on the measures' financial flows become available with the model-specific suitable detail.
- Fill data gaps through clear and transparent assumptions declared in your report.
- Classify the CGE model components and specify the model structures according to the RDP measures' priorities.
- Build systematic data bases specific to the model construction and the measures' financial flows.

Don'ts

- Forget to check for data availability before deciding which method to apply.
- Omit contacting the rural development experts, and getting assistance for economic structures and interpretation of findings.

4.8   EU 2020 Strategy and Innovation

4.8.1   Dos and don'ts (EU2020)

Dos

- Consult MAs for any monitoring data on headline targets (not only MAs for RDP, but also other operational programmes which relate to headline targets).
- Design with great attention to detail the questions/issues to be addressed prior to the screening of measures
- Consult RD experts for advice on the structure and content of the survey.
- Select comparable target groups for the survey.
- Check Eurostat definitions, indicators/data to ensure comparability of information between surveys and Eurostat data.
- Explore the existence of georeferenced data which may be of great help to assess the RDP contributions to headline targets.
- Establish synergies with evaluators working on CEQs that address headline targets, e.g., CEQ 24, CEQ 28, etc.

Don'ts

- Spend time on screening all RDP measures, focus on the ones depicted in the intervention logic of each headline target.
- Forget to use all available information, e.g. the calculation of common impact indicators and additional indicators, if used for assessing impacts or answering CEQs - Do not reinvent the wheel!

4.8.2   Dos and don'ts (Innovation)

Dos

- Consider innovation efforts and trends in the RDP territory as the baseline for the assessment of RDP innovation potential.
- Examine the innovation potential of all RDP measures / sub-measures, not only those which are primarily designed for this purpose.
- Check where the innovation is directly encouraged by the project selection criteria.

Don'ts

- Forget to consider the project selection criteria as the starting point to define innovation for your RDP territory.

4.9 Overview of indicators used in answering CEQ 22 - 26

Table 15. Overview of CAP impact, complementary result indicators and additional indicators suggested to answer the CEQ linked to the EU 2020 strategy headline targets

EU 2020 Strategy headline target	Common evaluation question for RDP 2014-2020	Suggested CAP impact indicators	Additional indicators
75% of the population aged 20-64 should be employed, corresponding to the EU2020 strategy priority of inclusive growth.	CEQ 22: To what extent has RDP contributed to achieving the EU 2020 headline target of raising the employment rate of the population aged 20-64 to at least 75%?	Rural employment rate (I.14).	Indicator related to the EU 2020 headline target: Employment rate of the population aged 20-64.
3% of EU's GDP should be invested in R&D, corresponding to the EU 2020 strategy priority of smart growth.	CEQ 23: To what extent has the RDP contributed to achieving the EU 2020 headline target of investing 3% of EU's GDP in research and development and innovation?		RDP expenditure in R&D as a % of GDP. Gross domestic expenditure on R&D (GERD) relative to gross domestic product (GDP). RDP expenditures in R&D and innovation as a % of the total RDP expenditures. RDP expenditures in R&D and innovation as a % of the gross domestic R&D & innovation expenditures.
The '20/20/20' climate/energy targets should be met (including an increase to 30% of emission reduction if the conditions are right), corresponding to the EU 2020 strategy priority of sustainable growth.	CEQ 24: To what extent has RDP contributed to: Climate change mitigation and adaptation, Achieving the EU's 2020 headline target of reducing greenhouse gas emissions by at least 20% compared to 1990 levels, or by 30% if the conditions are right, Increasing the share of renewable energy in final energy consumption to 20%, Achieving 20% increase in energy efficiency?	Emissions from agriculture (I.07). Increase in efficiency of energy use in agriculture and food processing in RDP supported projects (FA 5B - Complementary result indicator R14). Renewable energy produced from supported projects (FA 5C - Complementary result indicator R15). Reduced emission of methane and nitrous dioxide (FA 5D - Complementary result indicator). Reduced ammonia emissions (FA 5D - Complementary result indicator R19.	Indicators related to the EU 2020 headline target: % of GHG emissions as compared to 1990 levels, Share (%) of renewable energy in final energy consumption, % increase in energy efficiency, Ammonia emissions from agriculture GNB-N.
20 million less people should be at risk of poverty, corresponding also to the EU2020 strategy priority of inclusive growth.	CEQ 25: To what extent has the RDP contributed to achieving the EU 2020 headline target of reducing the number of Europeans living below the national poverty line?	Degree of rural poverty (I.15).	Number of People at risk of poverty or social exclusion.

89

EU 2020 Strategy headline target	Common evaluation question for RDP 2014-2020	Suggested CAP impact indicators	Additional indicators
Halting the loss of biodiversity and the degradation of ecosystem services in the EU by 2020, and restoring them in so far as feasible, while stepping up the EU contribution to averting global biodiversity loss, corresponding to the EU2020 Biodiversity strategy.	CEQ 26: To what extent has the RDP contributed to improving the environment and to achieving the EU Biodiversity strategy target of halting the loss of biodiversity and the degradation of ecosystem services, and to restore them?	Farmland Bird Index (FBI) (I.08). High Nature Value (HNV) farming (I.09). Water abstraction in agriculture (I.10). Water quality (I.11). Soil organic matter in arable land (I.12). Soil erosion by water (I.13). Ammonia emissions from agriculture (I.07).	Additional information on ecosystem services. Number of flora and fauna species on contracted land. Number of farmland bird individuals. Singing males of corncrakes (example of individual bird species indicator). Bumblebee indicator. EU Biodiversity Indicators linked to Target 3A – Agriculture and 3B – Forestry.

4.10   Glossary

This glossary contains only selected terms, that have not already been defined by DG AGRI's Technical Handbook for the Common Monitoring and Evaluation Framework (June 2017). For further terms please also refer to Evaluation Helpdesk (2014) Guidelines for the ex post evaluation of 2007- 2013 RDPs as well as to the updated Glossary "Key terms related to the evaluation of Rural Development Programmes 2014-2020" published at the webpage of the Evaluation Helpdesk.

Allocative efficiency

Allocative efficiency, also referred to as Pareto efficiency, occurs when resources are so allocated that it is not possible to make anyone better off without making someone else worse off.

Source: OECD Glossary of statistical terms.

Bottom-up evaluation

Set of techniques which allow to scale up the evaluation findings from the micro- to the macro-level (e.g. from farm to sector, from plot to the RDP area). For instance, these can be: GIS, satellite images or spatial analysis.

Source: Evaluation Helpdesk (2018)

Causal analysis

The analysis that attempts to establish a relationship of cause and effect between observed phenomena. In the case of evaluation, causal analysis attempts to establish a relationship of cause and effect between a public intervention and the changes (or lack thereof) observed in one or more outcomes of interest.

Source: EVALSED (2013) The resource for the evaluation of Socio - Economic Development. Evaluation guide.

Control Group

A group of study participants who have not been exposed to a particular treatment. The term is typically used in experimental designs with random assignment. A control group is closely related to a comparison group. However, whereas a comparison group is exposed to all the same conditions as the experimental group except for the variable that is being tested, the control group is not exposed to any condition.

Source: European Evaluation Helpdesk for Rural Development (2015)

Cost-effectiveness

Ability to provide sound evaluation findings whilst spending less money

Source: Evaluation Helpdesk (2018)

Credibility

Ability of the method to generate findings which can be trusted by stakeholders, for example the method demonstrates the causality, isolate programme effects from other factors, and eliminate the selection bias.

Source: Evaluation Helpdesk (2018)

Cut-off score

Especially used in Regression Discontinuity Design (RDD), a cut-off score is a predetermined threshold established to create the treated group, which includes all units at or above the threshold, as well as to create the comparison group, which includes all units below the threshold. The threshold is usually specified in terms of the size of some known relevant variable.

Disposable income

Disposable income includes all income from work (employee wages and earnings from self-employment); private income from investment and property; transfers between households; all social transfers received in cash including old-age pensions.

Evaluation approach

An evaluation approach is a way of conducting an evaluation. It covers its conceptualisation (purpose, objectives, evaluation standards, decisions on methods and tools applied in a certain combination as linked to available and collected data and information) and practical implementation (applying methods and tools) to produce evidence on the effects of intervention and its achievements).

Macro- and micro-level consistency check

In the context of these Guidelines, micro- and macro-level consistency check is the assessment of the correspondence or coherence between the evaluation findings observed at micro- and macro-level.

Naïve evaluation approaches

Naïve evaluation approaches are based on techniques which attribute the whole changes observed in a given indicator to the programme or intervention, without applying robust counterfactual analysis to exclude the confounding factors. These include: Before/After estimator, 'with' vs. 'without' approach, or comparison with population's average.

Practicability

Extent to which the method can be applied without adverse consequences (e.g. ethical) given the available data, resources, time.

Rigour

Ability to produce exact findings. Rigorous evaluation requires first of all to be able to rely on a causal analysis. Rigour in causal attribution of the applied quantitative evaluation method (part of an overall evaluation design) comes very close to the ideal, i.e. experimental design.

Reliability

Quality of the collection of evaluation data when the protocol used makes it possible to produce similar information during repeated observations in identical conditions.

Robustness

Ability to produce findings which are stable and resilient to small but deliberate changes

Transparency

Transparency of an evaluation methodology requires that users know exactly its main elements, structure, parameters, rules and functional responses. A user can therefore monitor that they are followed. A valid estimate of the counterfactual should be based on clear and transparent assignment rules.

Source: Technical Handbook on the Monitoring and Evaluation Framework of the Common Agricultural Policy 2014 – 2020 (2017).

Unit of analysis

The smallest part of an organised system which is being analysed. The unit of analysis can be defined at the micro and macro level of assessment. For instance, the unit of analysis at micro level could be parcels or farms whereas at macro level it could be catchment or NUTS 3, as well as the entire RDP territory.

Source: Evaluation Helpdesk (2018)

Unit of measurement

Used to observe a phenomenon, change or variable, and to place it on a quantitative scale. A measurement unit allows for quantification. An elementary indicator is associated with a measurement unit and has only one dimension (e.g. 10km of motorway; number of training courses). Some measurement units are divisible and others not (e.g. 20.3 km were built; 30 trainees were qualified). Measurement units must be harmonised if indicators are to be comparable.

Source: Technical Handbook on the Monitoring and Evaluation Framework of the Common Agricultural Policy 2014 – 2020 (2017).

Validity

Accuracy, logical and factual soundness of method in depicting the reality without errors and the conclusions and decisions based on this depiction.

Source: Evaluation Helpdesk (2018)