ENVIEVAL

Development and application of new methodological frameworks for the evaluation of environmental impacts of rural development programmes in the EU

(Project Reference: 312071)

Area 2.1.4: Socioeconomic research and support to policies

KBBE.2012.1.4-08: Development and application of methodologies and tools, including indicators, for the assessment of environmental impacts of rural development programmes in the EU

WP3, WP4 and WP5

Internal report on logic model for a framework for counterfactual development, at micro and macro levels of environmental evaluations (Draft, Version November 2015)

Authors: ENVIEVAL consortium

This document was produced under the terms and conditions of Grant Agreement No. 312071 for the European Commission. It does not necessary reflect the view of the European Union and in no way anticipates the Commission's future policy in this area.

Table of Contents

Table of Figures

1 General concept and logic model

The main aims of the logic models are to assist evaluators to find a sound evaluation design for the task at hand and to help managing authorities to assess the feasibility of evaluation plans and/or the quality of evaluation results. The general concept of the logic models follows a nested approach with different layers for the design of counterfactuals, micro level and macro level evaluations. Figure 1.1 provides a simplified overview of the general structure of the logic model layers.

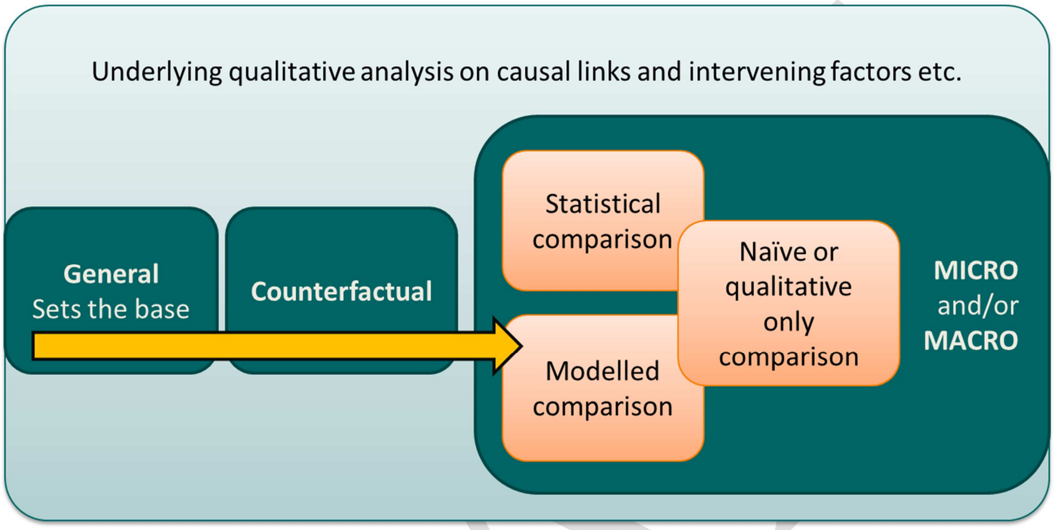

Figure 1.1: Simplified logic model flow

2 General workflow and description

The first general layer of the logic models provides an overview of the overall intervention logic and structure suggested to pursue in the evaluations of environmental RDP impacts (from the decision to evaluate specific measures or the whole programme to the integration or dis- and aggregation of micro and macro level results to a consistent net impact assessment).

Step 1.1 – following the CMES intervention logic

Evaluators are facing the tasks of assessing the environmental impacts of the programme and different relevant measures. Starting with the formal requirements and general intervention logic of the Common Monitoring and Evaluation System (CMES) evaluators need to select relevant measures of the RDP and evaluation questions for the environmental objective(s) they want to evaluate the measures / programme against. Then, output, result and impact indicators need to be selected and reviewed in the context of the available data (Figure 2.1)[1]. It is recommended to also

[1] Result indicators of the CMEF relevant for environmental aspects do not provide suitable proxy for the measurement of environmental changes and, consequently, for the evaluation of environmental impacts of the RD measures and programmes.

consider at this stage to what extent the available data will later on enable the coverage of unintended effects on the environment and indirect effects such as deadweight and leverage effects at micro level and substitution and displacement effects at macro level. The inclusion of indirect effects into the evaluation design requires sufficient available data for non-participants.

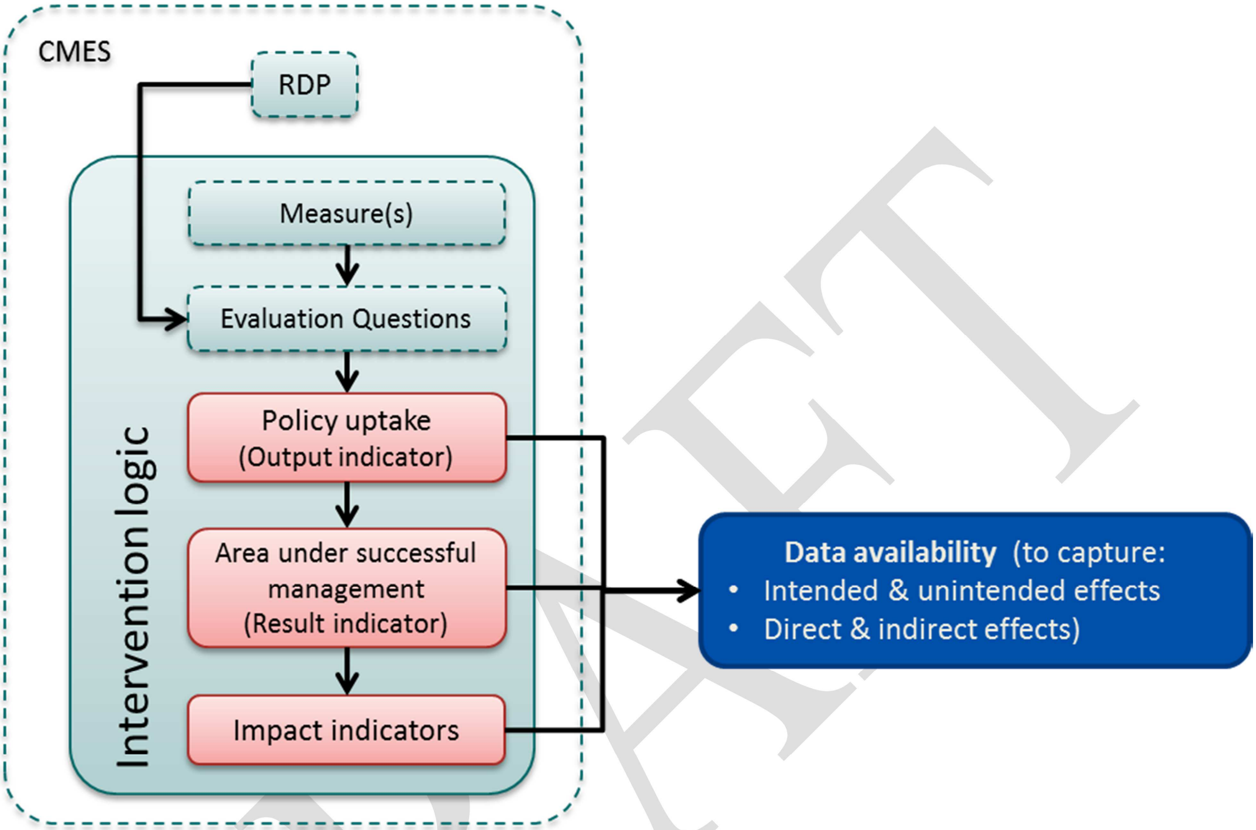

Figure 2.1: Step 1.1 - CMES requirements in the general logic model

Step 1.2 – selecting additional environmental indicators

While the CMES provides useful guidance on the general intervention logic, the number of environmental impact indicators is limited and depending on the public good and environmental objective against which the measure and programme is evaluated, it becomes necessary to identify and select more suitable indicators to quantify environmental changes and to establish robust causal relationships between the policy induced land management (or livestock management in the case of animal welfare) changes and measured environmental change. The suitability of the selected indicators needs to be reviewed in the context of their data requirements and the available environmental monitoring data Figure 2.2. As highlighted in Step 1.1, it is recommended to also review at this stage to what extent available environmental monitoring data cover non-participants and will later on enable the coverage of unintended effects on the environment as well as indirect effects such as deadweight effects at micro level and substitution effects at macro level.

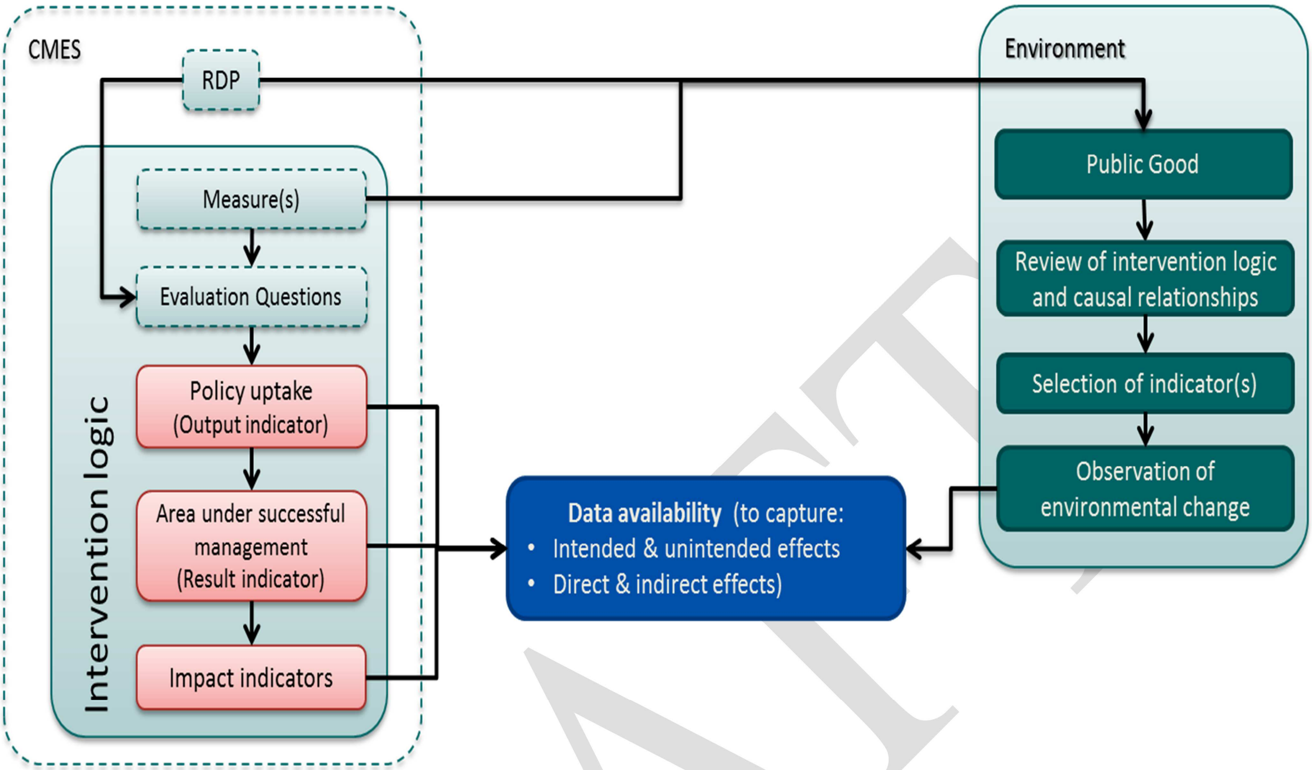

Figure 2.2: Step 1.2 - Selection of additional environmental indicators

Step 1.3 – defining units of analysis for micro and macro level evaluations

Depending on the public good and environmental objective, the selected indicators and available data as well as the level of analysis (micro or macro), a common unit applied to all used data needs to be defined for micro level and macro level evaluations (Figure 2.3). The unit of analysis can be defined as the "smallest part of an organized system" (parcels, farm as agro-ecosystem, landscape unit, ecological area, sub-catchment area, etc.). The unit of analysis refers to the unit of study for assessing functional contributions of a system under a specified metric and delimits the analysis and the comparison of the organized system. Furthermore, the units of analysis are characterized by homogeneous activities and allow solving the scale interdependencies which is an important aspect to be define for the logic model implementation. Examples of common units of analysis include farm (micro), catchment and regional units (macro).

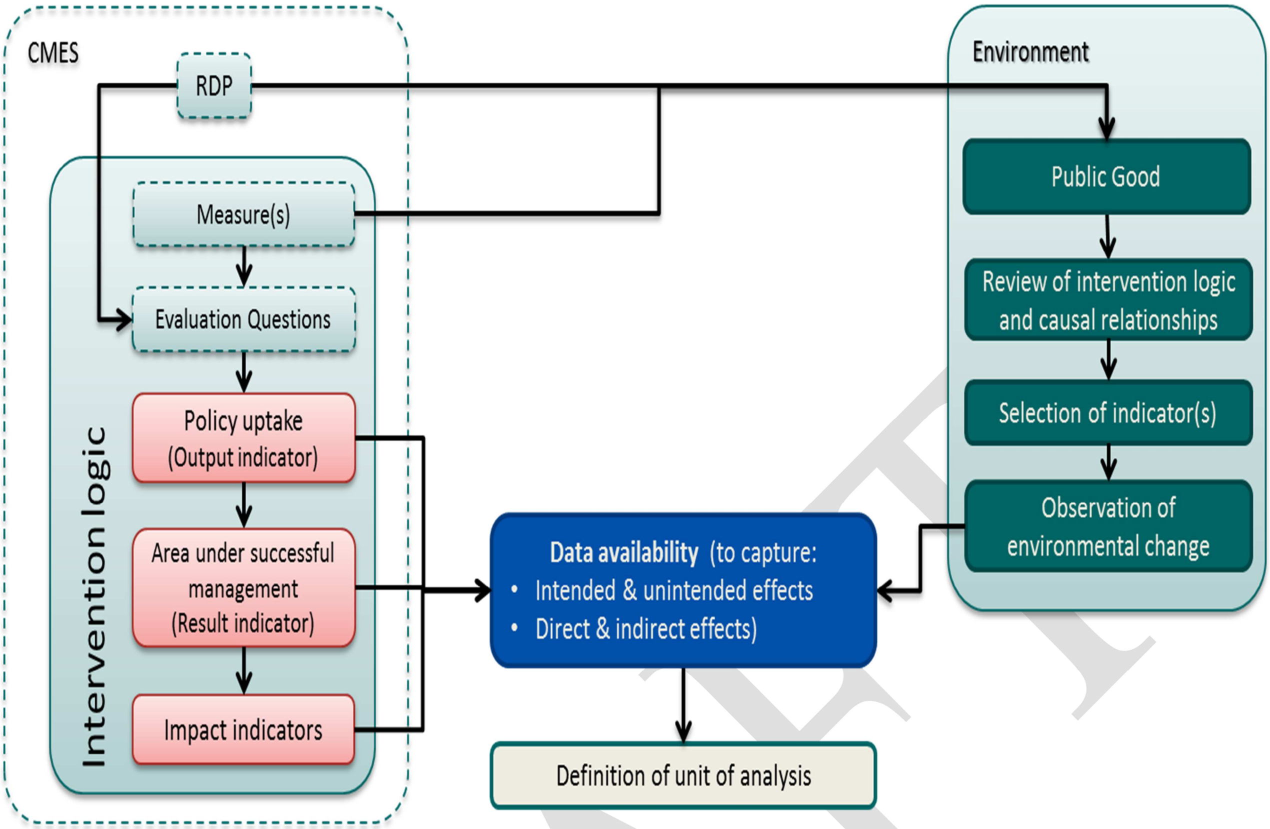

Figure 2.3: Step 1.3 - Definition of common functional units for micro and macro level evaluations

Steps 1.4 – Conceptual decision on counterfactual micro and / or macro level evaluations

Counterfactual based micro level evaluations are then designed and possible aggregation or upscaling of micro level data and results to macro level are reviewed. Alternatively, a separate counterfactual-based evaluation design is developed for macro level assessments. In either case, consistency checks between micro level and macro level results are required. The following layers of the logic models explain the different steps in designing and carrying out counterfactual based micro and macro level evaluations of environmental impacts of RDPs. Figure 2.4 shows the complete general logic model layer.

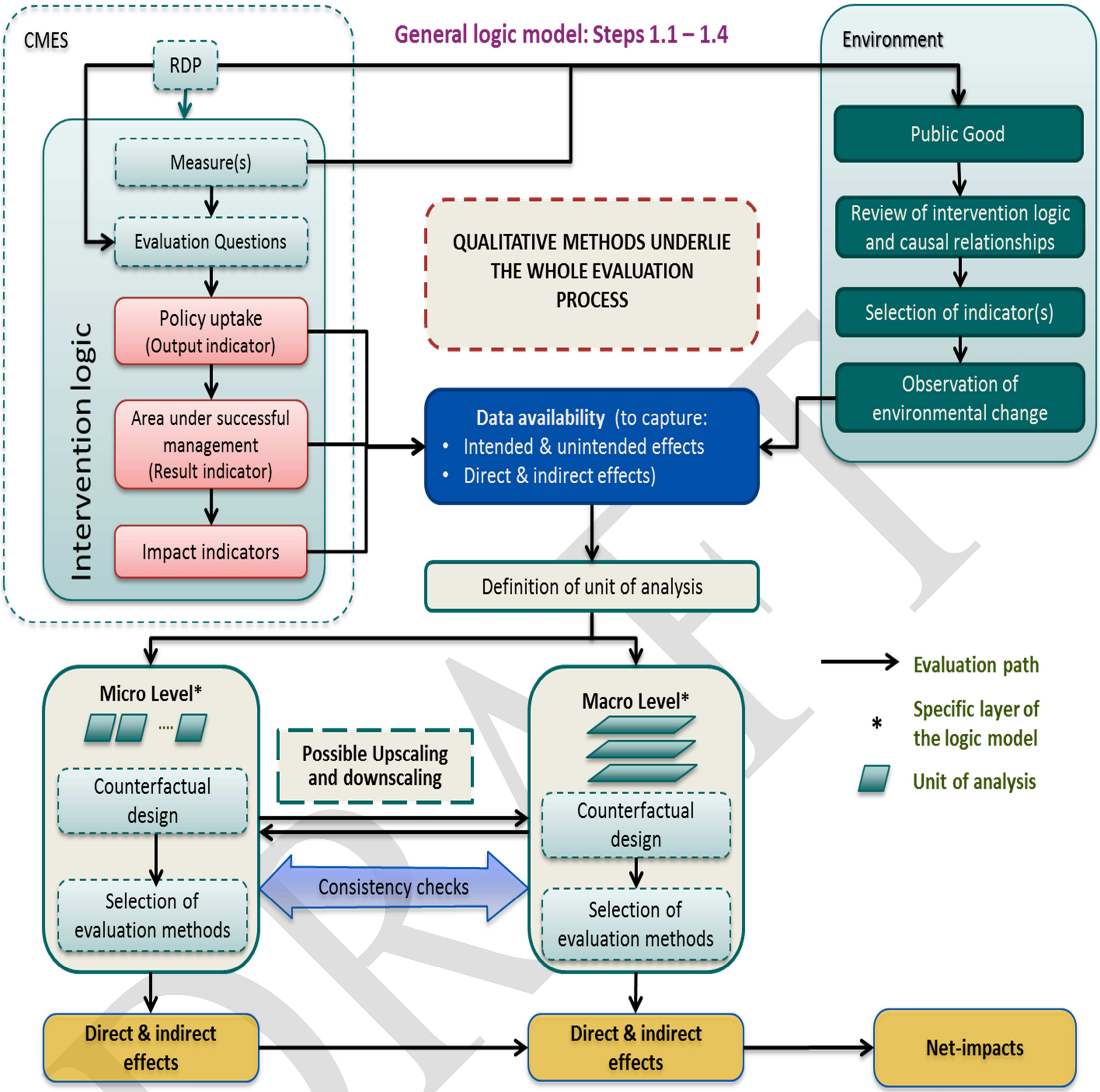

Figure 2.4: Step 1.4 - Counterfactual micro and macro level evaluations and net-impact assessment

3 Workflow and description of the counterfactual design

The following steps describe the workflow in going through the logic model to design counterfactuals. The structure is general and works for both micro and macro stages, providing a set of questions that need to be answered before deciding on the counterfactual design. Most importantly, the logic model highlights the importance of comparison groups. Comparison group formation is particularly important when there is self-selection to programme participation. As farmers are not randomly assigned as participants to the programme, comparing all programme participants and non-participants directly may suffer from sample selection bias[2] (see D3.1 for more information).

While the logic model considers comparison groups predominantly from data perspective, their identification is important even when data does not allow for their full individualization. The recognized comparison groups work as a basis for deciding the point of comparison, i.e. the counterfactual, in cases where the logic model guides to evaluation approaches that are less dependent on statistics.

Figure 3.1 shows the full logic model for designing counterfactuals. In the following text, the logic model is broken into smaller parts and explained separately.

Following this logic model, the evaluator should know the type of counterfactual she is able to construct with the available data on a general level. This consideration is then subjected to the micro and macro level specific logic models to further refine the evaluation possibilities. Micro and macro level specifications are initially decided in the general logic model, where available data is a driver on the decision on the level of analysis (micro, macro or both). Thus micro and macro levels have their separate considerations on counterfactual development.

We note that due to the plurality of different evaluation methods and the need to keep the logic model as uncomplicated as possible, the final suitability of the available data for each method, especially for the statistical methods, must always be assessed case-by-case with experts. With small sample sizes it is possible that the logic model suggests a statistical approach that cannot be conducted with the data. Each comparison group in statistical analysis should have at least 30 observations of the functional units. Further, we have assumed in this logic model that the data are spatially and temporally synchronized with the functional units and the programme period. In cases

[2] In rare cases naturally random participation or experiments can be found. With voluntary participation to programmes, evaluation planning cannot rely on such occurrences. Thus the logic model stresses the comparability of groups to form a counterfactual.

where this assumption does not hold, the evaluator needs to qualitatively assess the risks and magnitude of bias in the evaluation results.

Further, the logic model works best for cases with a single indicator. This translates to cases where a measure or a group of measures aim for the same environmental outcome. Programme-level evaluations are often too abstract to be evaluated with a single indicator and, hence, a single evaluation method. Rather, the logic model offers tools with which to construct a sound counterfactual-based evaluation that may include a number of approaches in programme-level impact evaluation. Thus for a programme level evaluation, the evaluators should use the logic model to collect and compare the different evaluation possibilities and generate a concise evaluation plan that uses similar counterfactual scenarios across the line. In practice, the last requirement can be challenging. The logic model, however, offers a structured identification plan for counterfactual scenarios enabling discussion on the impact of different counterfactual scenarios to the programme level evaluation

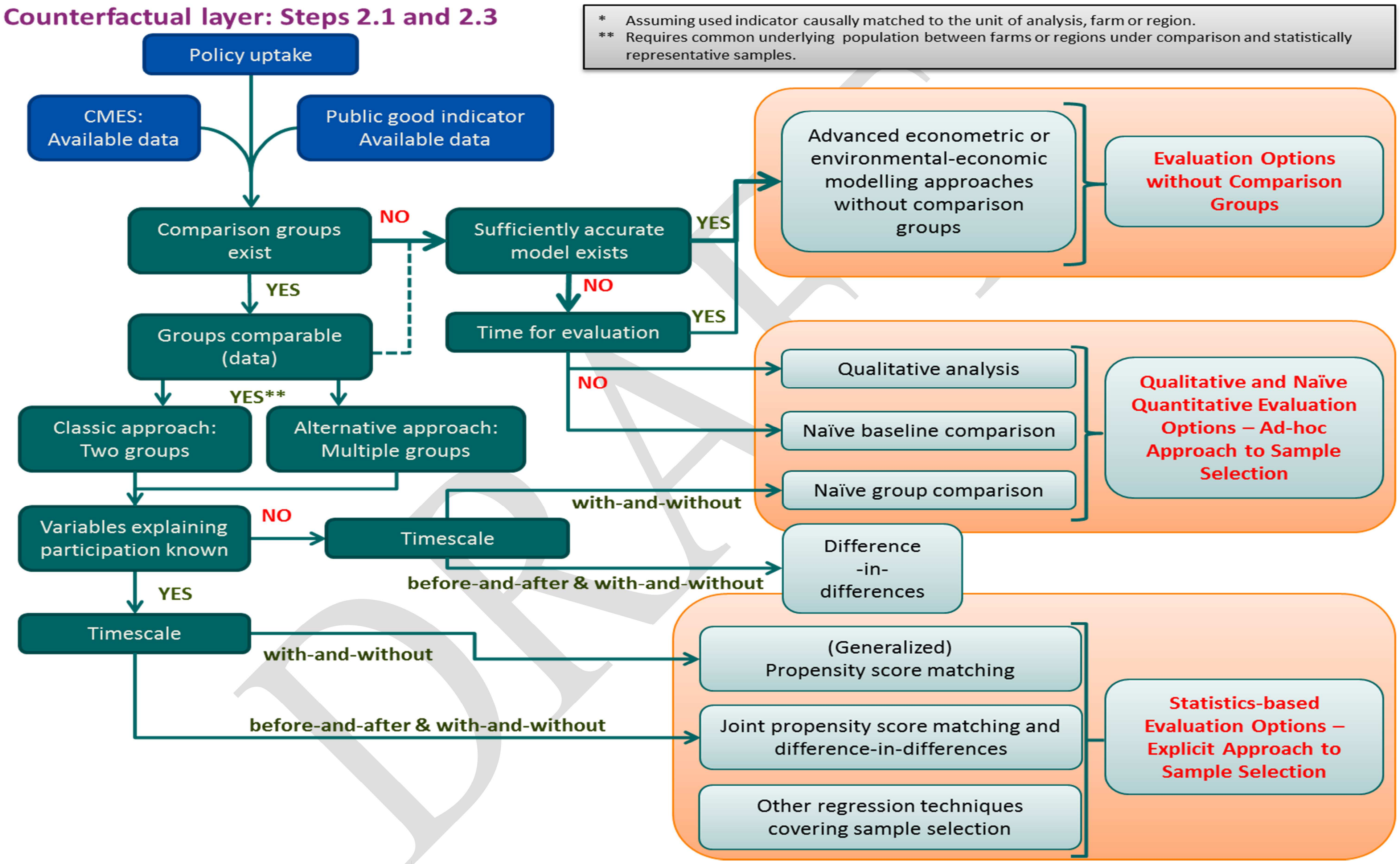

Figure 3.1: Workflow and description of the counterfactual design

Step 2.1

At this stage (Figure 3.2: Input to the counterfactual logic model) the counterfactual logic model starts with the data availability found at the general logic model. CMES data provides data for the official common evaluation question(s) and indicators, whereas public goods indicators are case specific, when the CMES indicators do not suffice to measure the environmental impact at the required level. Policy uptake determines the possible actual comparison groups in later stages, initially indicating the farms that have partaken in the evaluated measure and those that have not. At this stage the logic model is not strict on the micro or macro level of analysis - the methods of comparison do not largely disagree with scale. However, the micro and macro specific logic models further refine the possible approaches to counterfactuals.

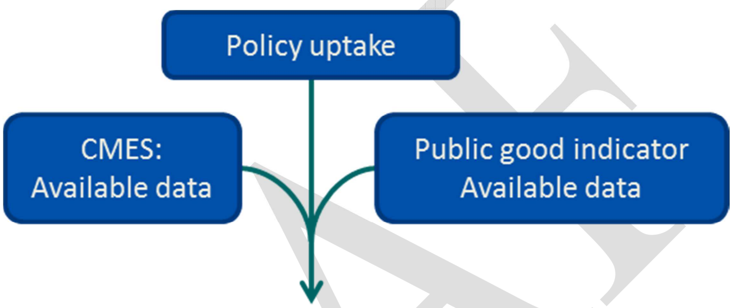

Figure 3.2: Input to the counterfactual logic model

Step 2.2

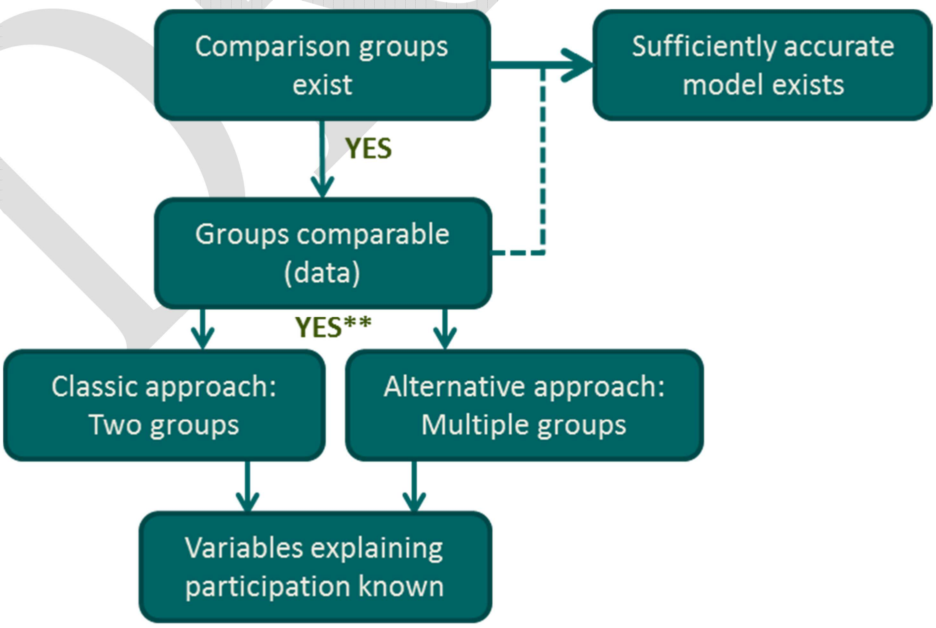

Figure 3.3: Comparison group definition stage

At this stage (Figure 3.2) the data from the previous stage are collected together. Within these data it is then searched if natural comparison groups arise, e.g. participants and non-participants (Figure 3.3). The quality and quantity of the data should be then further assessed on how they allow the actual construction of comparison groups.

As statistical methods might have different subroutines for more than two comparison groups the parallel stage is identified with multiple groups. First two stages of comparison group formation include exits for other options on forming the groups of comparison. Note that the logic model appears to prefer quantitative approaches over qualitative. However, careful qualitative assessment should underlie each stage of the logic model[3], e.g. identification and choices of suitable comparison groups for assessment is subject to qualitative approaches the more complex the setting gets. Below is a helping matrix from D3.2 explaining the number of possible comparison groups under different intervening effects. The matrix is not exhaustive, as intervening effects are many in form. The matrix serves to show the increasing complexity for statistical approaches with multiple effects. Qualitative approaches may be needed to describe the possible severity of each effect to and narrow down the number of comparison groups to tractable levels.

Participation status in evaluation period	Eligibility rules exist for participation	Deadweight (internal)	Deadweight (external)	Minimum number of groups
Only participants/ non-participants (2)	All eligible (x1)	None (x1)	None (x1)	2
			Historically significant outside pressure at min. one area (+2)	4
		Previous participation status affects environmental effects or participation probability (x2)	None (x1)	4
			Historically significant outside pressure at min. one area (+2)	6
	Some non-participants ineligible or in a queue to participate (+1)	None (x1)	None (x1)	3
			Historically significant outside pressure at min. one area (+2)	5
		Previous participation status affects environmental effects or participation	None (x1)	6
			Historically significant outside pressure at min.	8

[3] This qualitative assessment is important especially in cases where the intervention logic of the measure or policy and data collection have not been well linked in practice.

Participation status in evaluation period	Eligibility rules exist for participation	Deadweight (internal)	Deadweight (external)	Minimum number of groups
		probability (x2)	one area (+2)	
Participants/ non-participants, also drop outs and/or late joiners (3/4)	All eligible (x1)	None (x1)	None (x1)	3/4
			Historically significant outside pressure at min. one area (+2)	5/6
		Previous participation status affects environmental effects or participation probability (x2)	None (x1)	6/8
			Historically significant outside pressure at min. one area (+2)	8/10
	Some non-participants ineligible or in a queue to participate (+1)	None (x1)	None (x1)	4/5
			Historically significant outside pressure at min. one area (+2)	6/7
		Previous participation status affects environmental effects or participation probability (x2)	None (x1)	8/10
			Historically significant outside pressure at min. one area (+2)	10/12
No non-participants (1)	All eligible (x1)	None (x1)	None (x1)	no statistical comparison possible
		Previous participation status affects environmental effects (x2)	None (x1)	2, note counterfactual is not for inaction
			Historically significant outside pressure at min. one area (+2)	4, note counterfactual is not for inaction
No non-participants but late joiners (2)	All eligible, queues to participate (x1)	None (x1)	None (x1)	2, note counterfactual for partial measure participation
			Historically significant outside pressure at min. one area (+2)	4, note counterfactual for partial measure participation
		Previous participation status affects environmental effects (x2)	None (x1)	4, note counterfactual for partial measure participation
			Historically significant outside pressure at min. one area (+2)	6, note counterfactual for partial measure participation

Step 2.3

At this stage (Figure 3.3.) the number of possible comparison groups has been decided. Then the data are searched for variables that may statistically explain participation to the measure. The evaluator addresses specifically sample selection issues, where comparison groups may differ by population type due to different underlying qualities. Variables that explain participation to the evaluated measure or policy are case-specific and depend on the functional unit (e.g. regional vs. farm uptake of a measure). These variables should the very least include factors of functional units that are targeted in the measure or policy. For example if the policy targets cereal producers below certain income level in high risk erosion areas, at least farmer income, production type and having fields in high risk erosion areas should be known for all comparison groups. For regional analysis this level of analysis may be difficult. The next step is to assess the timescale of data. The timescale can be based on cross-section data (with-and-without), i.e. data for only at the end of the evaluation period, or to data gathered both at the beginning and end of the evaluation period[4]. Each step brings a new method for statistical analysis. The exit from having no variables explaining participation and only ex-post data available leads to naïve approaches in evaluation.

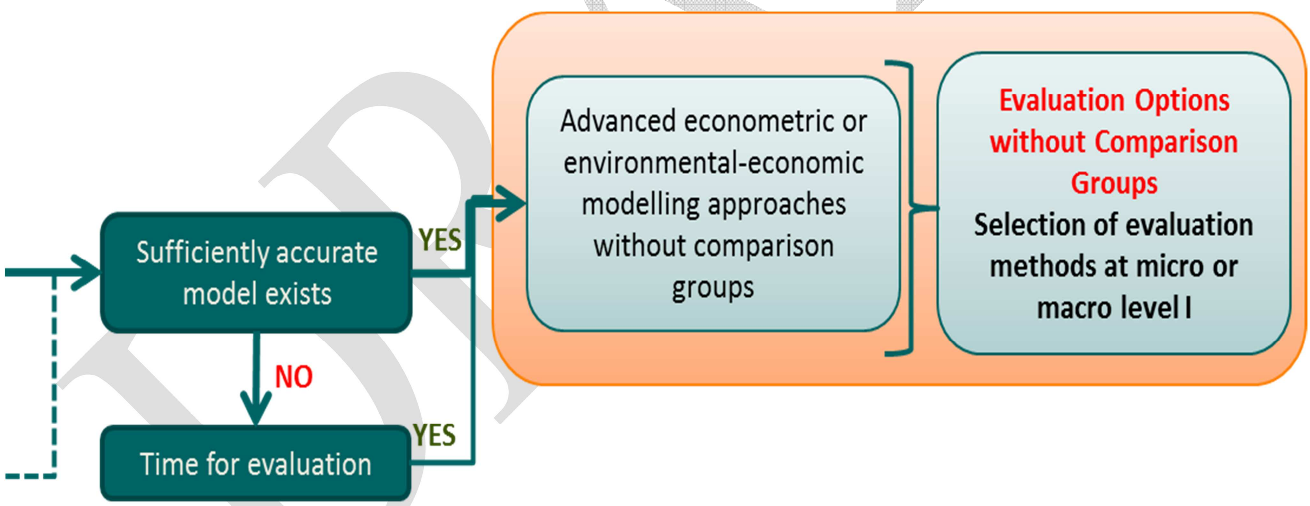

Figure 3.4: Step 2.3a Choice of long run evaluation options without comparison groups for counterfactual construction

In Figure 3.4 if a sufficiently accurate model does not exist or non-participants are in no way comparable to the participants (i.e. comparison groups cannot be separated in data), if there is time and resources we recommend building a model that enables a consistent counterfactual study of effects also in future evaluations. Such an approach enables a self-updating system for evaluation

[4] and intermediate periods if there are late-joiners and drop-outs in program participation

with new data and possibilities to build on earlier modelling work. It should be noted, that the evaluator may also choose this option over statistical approaches if the model approach is the cost-efficient solution. Further, the counterfactual still needs to be decided, which requires the identification and decision of a policy scenario. A number of scenarios are possible in modelling, including a simple business-as-usual scenario without the evaluated measure. Such counterfactual scenarios could be decided to follow the most credible path should the measure not have been implemented. The evaluator needs to decide if other intervening policies or measures should be taken into account in the counterfactual scenario.

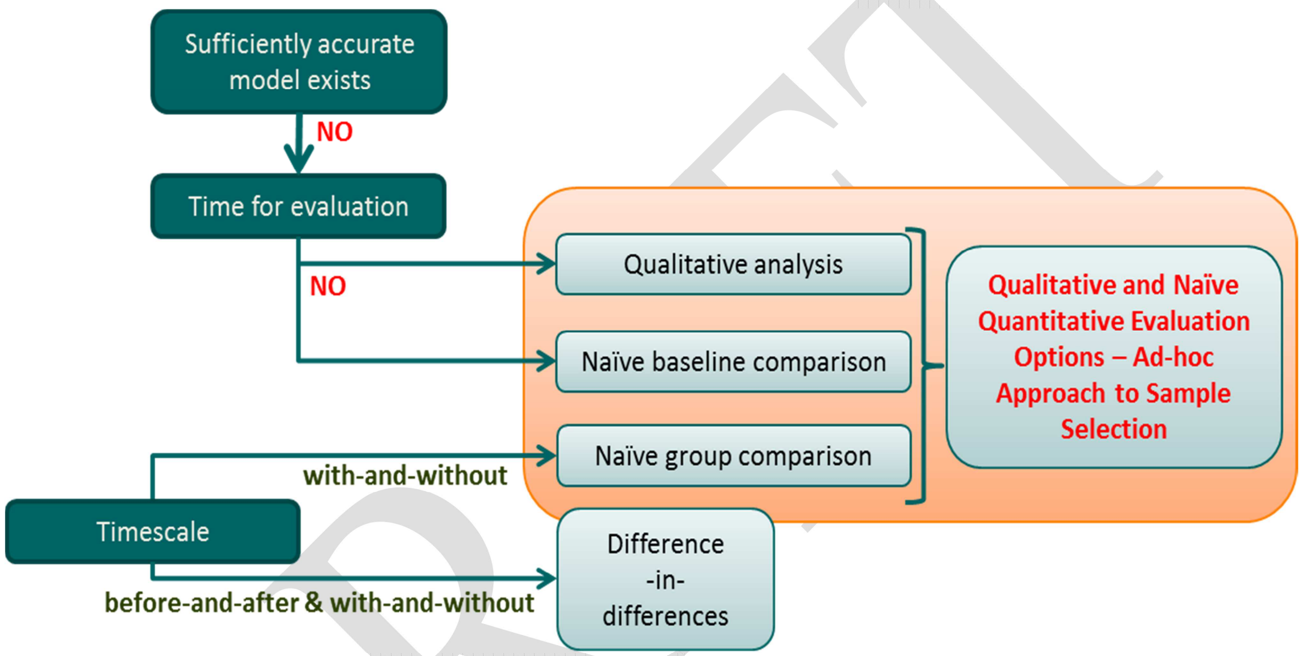

Figure 3.4: Step 2.3b Choice of naïve statistical comparisons or qualitative analysis of counterfactuals

In Figure 3.5, if case time and resources are lacking for modelling approaches, or a statistical separation of comparison groups is not possible, the set of naïve counterfactuals is possible for the evaluators for designing a counterfactual. If the data on comparison groups does not include additional information, i.e. variables, that would explain why farms would participate or self-select to the evaluated measure, the next step is to determine if the data covers multiple points in time. If there is sufficient data for participants and non-participants for both before and after the evaluation period, the evaluator can resort to the difference-in-difference (DiD) method family. The DiD can be considered either a naïve quantitative or a statistics-based (see Figure 3.6) approach depending on the underlying assumption on data quality. If the evaluator believes the participant and non-participant groups would have different time trends for the environmental indicator development, the evaluator can opt for an ad-hoc sample selection correction in the data. For example, in the Italian climate stability case study, the evaluator chose the sample for analysis based on experience to compare neighbouring participants and non-participants. In the absence of dissimilar time trends across comparison groups, the DiD is no longer considered as a naïve quantitative approach to sample selection.

When data is available only for cross-section at the end of the evaluation period, the evaluator has to conduct naïve quantitative analysis comparing the environmental indicator level with different comparison groups. The typical counterfactual would be provided the comparison group that has not participated in the measure. Following the naïve estimation of the effects, the evaluator should make a careful qualitative assessment on what internal and external factors are likely to affect the evaluation results. Is it likely that if there was self-selection of the measure, are the participants similar to non-participants in other respects? If there is self-selection, is the bias in the evaluation results likely an over- or underestimate, and how severe is the bias likely to be?

If there is no existing model, or time to make one, that could construct the counterfactual, the evaluator cannot make a fully data-based assumption of the counterfactual. In case of a missing comparison group but with some level of data for participants, the evaluator may make a naïve baseline assumption based on qualitative analysis. For example, expert opinions can be used to determine if a trend exists in the development of the environmental indicator. The counterfactual can be based on this trend and then contrasted to the observed level of the indicator. This way the evaluator essentially decides the level of environmental impact from the measure. Therefore the decision, its arguments and how it affects the counterfactual should be well and explicitly documented with qualitative sensitivity analysis if possible.

If the evaluator cannot use data for quantitative analysis for any reason, e.g. lack of data, poor quality of data, poor causal link between the measure and the impact, the only way to evaluate the impact of the measure is through qualitative analysis. There are numerous methods in the qualitative analysis literature to develop a counterfactual, but they often rely on different constructs by expert panels. Qualitative analysis requires careful discussion in the evaluation report showing the thought structure behind the analysis. Care should be taken to keep the results understandable. It is clear that qualitative analysis has a very strong foothold as a part of the evaluation procedure in step 4 as quantitative data cannot be fully used to back the counterfactual development.

Evaluators have been often forced to resort to ad-hoc approaches due to lack of sufficient data. In future evaluations, if data collection is designed from the beginning to support certain methods, modelling based or elaborated statistics based analyses are possible. Ideally, qualitative approaches would then be integrated in mixed method approaches in combination with advanced quasi-experimental methods (Figure 3.6).

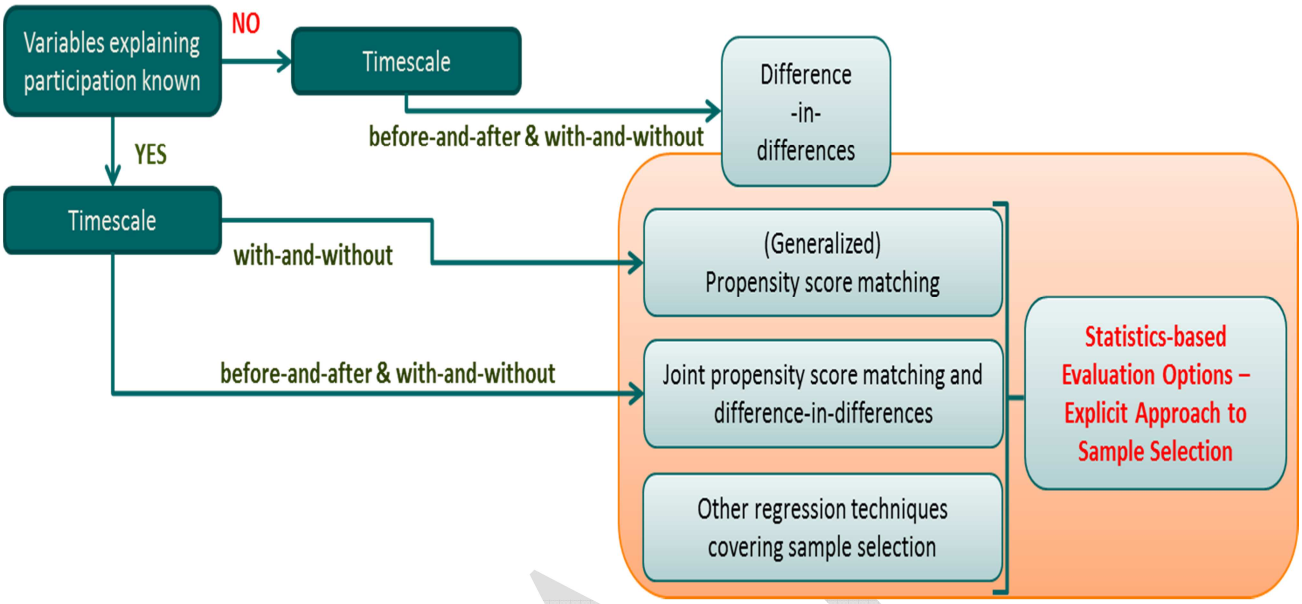

Figure 3.5: Step 2.3c Choice of elaborate statistics-based evaluation for counterfactual construction

Stages in Figure 3.6 involve the exits from the flow that lead to statistical approaches taking sample selection issues into account. There are two critical stages that determine whether statistics-based evaluation options can be used. First, variables explaining the participation in the measure should be known for all comparison groups evaluated. At a minimum the data should cover those participating and those not participating in the measure for the whole evaluation period. The variables explaining participation to the measure are case specific, but include typically the type of produce, size of production, and factors that are targeted by the measure.

Further, the temporal scale of the data should be examined. If data covers the moment before the implementation of the evaluated measure and the end of the evaluation period, the evaluator can resort to statistical methods even if variables explaining participation are unknown. The approach in such a case would be a difference-in-differences comparison of the changes in the environmental indicator during the evaluation period between participant and non-participant groups. In such a case the evaluator assumes that the trend in the environmental indicator is the same between the two groups.

If the evaluator can identify the variables explaining participation, but data exists only for the time after the evaluation period, the evaluator can use the propensity score matching method to construct the counterfactual. The propensity score matching method does, however, greatly benefit from having data on the farms before the programme period on variables explaining participation, as participation itself may cause these variables to change over the programme period.

The propensity score matching method compares participant and non-participant farms that have similar propensity to participate in the measure, thus alleviating sample selection bias of a naïve group comparison. There must be enough data on similar farms in both groups for the statistical model to work. If the participants and non-participants come from highly different populations, their comparison may not make sense using purely statistical methods. The generalized propensity score matching method can accommodate more than two comparison groups (and participation (treatment) levels representable by a continuous variable), but can be more challenging for the evaluator in terms of data requirements and methodological expertise.

The best opportunities for data-based statistical approaches in developing the counterfactual arise when the data allows a comparison of participant and non-participant groups (i.e. with-and-without) and the situation before-and-after the evaluation period. This is possible in cases where large scale farm monitoring data can be joined with equally precise environmental monitoring data. A clear candidate method of analysis is the joint propensity score matching and difference-in-difference method that enjoys the benefits of both methodologies.

We do note that the statistical and econometric toolbox offers a number of other approaches that tackle sample selection issues. These methods are explored to some extent in the ENVIEVAL Report D3.1 Review of counterfactual methods. In many cases these statistical methods are tailored to the case, and require statistical and econometric expertise from the evaluator. However, if the evaluation question and data remain uniform over time, repeating the evaluation becomes easier with experience. In addition, qualitative analyses and peer-review processes of elaborate statistical evaluations are possible for validation purposes.

It is possible to use the elaborate statistical methods for any two (or more) comparison groups. If a clear non-participant group data would not suffice for statistical analysis, but a reasonably close – an almost non-participant – group is identified and has the required data, it can be used to develop a counterfactual that resembles non-participation.

4 Workflow and description of the micro level logic model

The workflow for the micro level logic models leads to different methods which will contribute, through the integration of micro and macro level results, to a consistent net impact assessment. For each of the three possible counterfactual designs an individual micro level logic model has been created. The initial two phases of the workflow for these three logic models are the same and only from the third phase each of the counterfactual approaches leads to different micro level methods which will discussed separately later.

Step 3.1 - Definition of Functional Units and Indicators

At this stage the micro logic model starts with the general layer on the data availability for all the three counterfactual approaches (figure 4.1). To a large extent data availability determines the type of units of analysis that can be used in the evaluation process and it provides information on the suitable indicators to be developed according to the functional unit level and scale.

The selected indicators will subsequently define the specific scale at micro level. Available data and linkages to micro level results are considered for the identification of a unit of analysis, which leads to consistent indicators for the assessment.

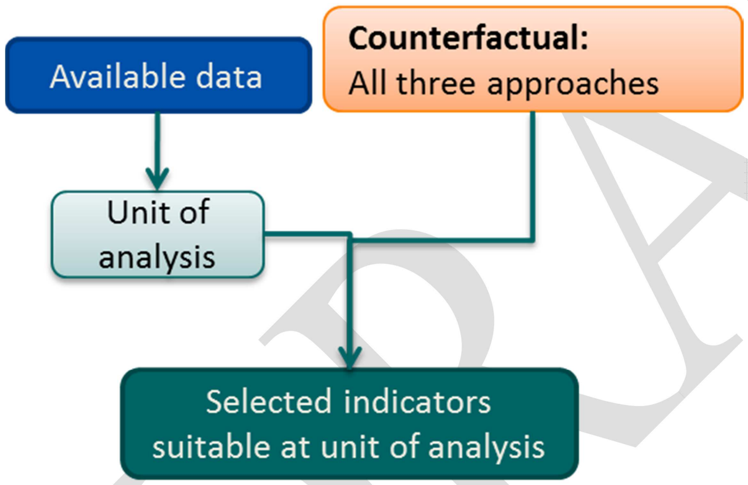

Figure 4.1: Definition of unit of analysis and indicators (Step 3.1 - micro level logic model)

Step 3.2 - Assessment of data quality

In the second phase the assessments of the quality of data[5] have been carried out in order to check if the amount and characteristics of data are appropriate to implement one of the models available for the impact evaluation (figure 4.2). The limited data quality often affects the applicability of the models for the environmental impacts assessment steering toward lack of consistent, robust or representative results. For this reason an essential point in the workflow is the identification of potential bottlenecks, due to poor data quality, that can make inadequate the calculation of the

[5] Among the various categories of attributes of data quality, the most commonly attributes included are: accuracy, correctness, currency, completeness and relevance.

selected indicators through one of the models. Furthermore in the case of environmental assessment the availability of spatially explicit data could make the difference between a rather descriptive survey and a more in-depth scientifically sound analysis. Starting from the selected indicators, a first check of the suitability of the data could require new primary data collected through statistical samplings. A second check could be needed to obtain sufficiently accurate data, possibly spatially explicit. In this case additional data and/or particular data processing are required to improve data quality.

Sometimes poor data quality is also due to the lack of access to administrative and statistical databases because of privacy regulations or ill-coordinated efforts to collect information for monitoring purposes. The minimal requirements for suitable data sources should be causally linked to each other and frequently monitored. For this reason, the use of qualitative approaches (common sense) is quite frequent, not only for the lack of data or of financial resources for creating new databases but also for the difficulties encountered by evaluators using complex methodologies that could guarantee a good outcome from the evaluation process. This knowledge gap has to be taken into account during the selection of a specific model which needs suitable data.

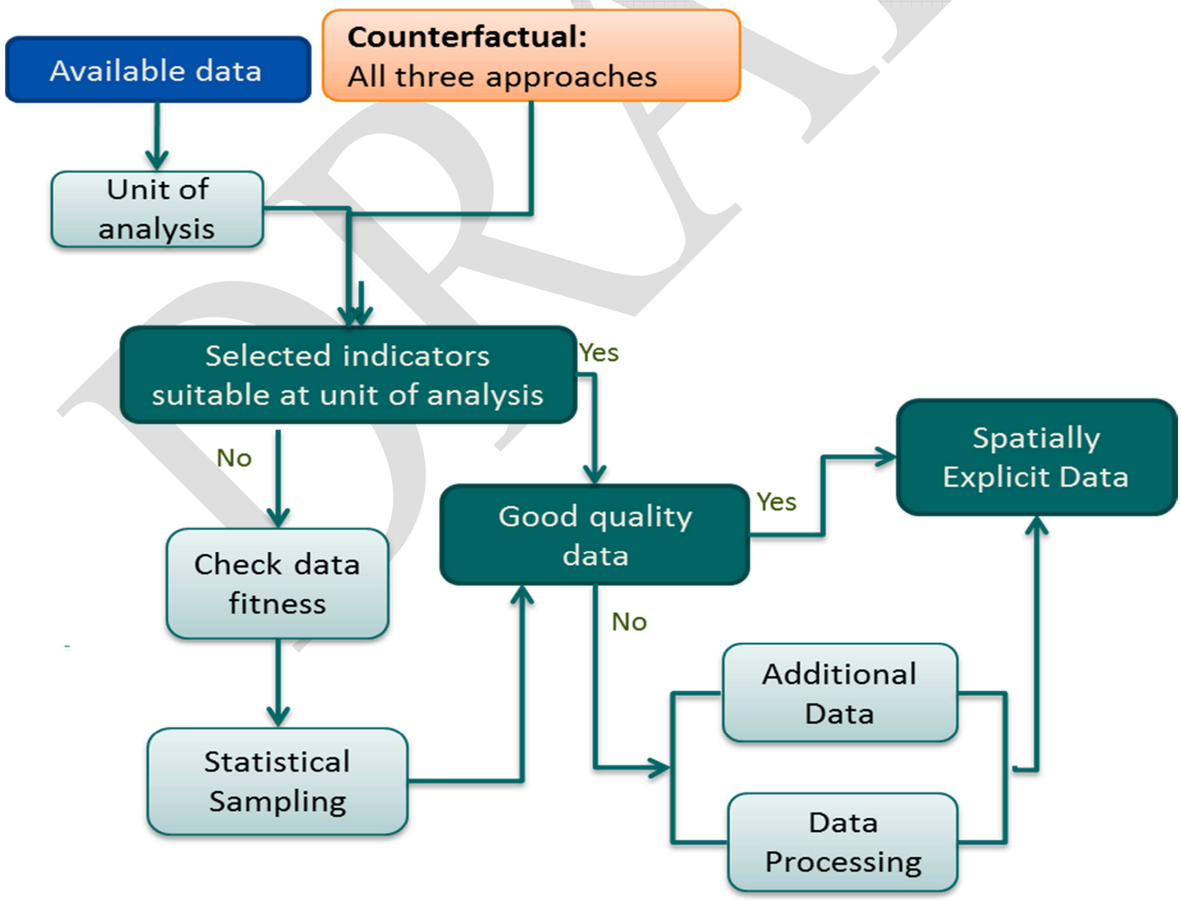

Figure 4.2: Assessment of data quality (Step 3.2 - micro level logic model)

After the data are collated and verified, selected models on the basis of the three counterfactual approaches have been identifying.

Step 3.3a - Long Run Evaluation Options w/o Comparison Groups

Without the control group (comparison group) of non-participant it is not possible to use the statistical approach in the counterfactual analysis. In this case three options can be defined:

i) conducting an intermediate counterfactual analysis between different participant groups (e.g. participants and late joiners),

ii) using similar non-eligible farms/regions to represent non-participants (regression discontinuity method), and

iii) comparing farms participating and those in queue together (pipeline method).

In presence of the above categories the selected methods are the structural model, integrated models and agent-based models, according to availability of spatially explicit data. Without spatial data **structural models** are more appropriate at the micro level. These models are defined by a mathematical approach to study the link between cause-effect relationships. More precisely the method builds a framework to interpreting policy effects due to specific interrelationships among endogenous variables and exogenous variables or factors without necessity of comparison group. This allows capturing the effects of specific environmental policies at micro level, due to focus of cause-effect relationships. In general the structural model can be used to estimate unobserved or behavioural parameter.

In case of availability of spatial data, the methods selected are the integrated models and the agent-based approaches. **Integrated models** allow addressing more holistically agri-environment evaluation questions, in particular at the farm scale and its sub-sets, such as cropped areas or parcels. In fact, this is the level for which farmers allocate available land and resources to the various tasks in their production systems. Integrated models are therefore able to shed light on the environmental components allowing evaluation of specific programmes. The environmental impacts of these changes can be estimated introducing linkages with bio-physical models at farm scale.

To date, researchers use farm-level decision models to assess behaviours and changes with **Agent-Based Modelling** (ABMs) approaches in ex-ante evaluation exercises. These approaches allow the coupling of environmental models and the social systems embedded in them. In this way the role of social interactions of adaptive, disaggregated (micro-level) human decision making processes in environmental management can be modelled. In short, the development and use of ABMs for ecosystem management allows considering ecological complexity. It is possible to identify the role of individuals and to analyse in more depth and more effectively the different forms of organisation (spatial, networks, hierarchies) and interactions among different organisational levels.

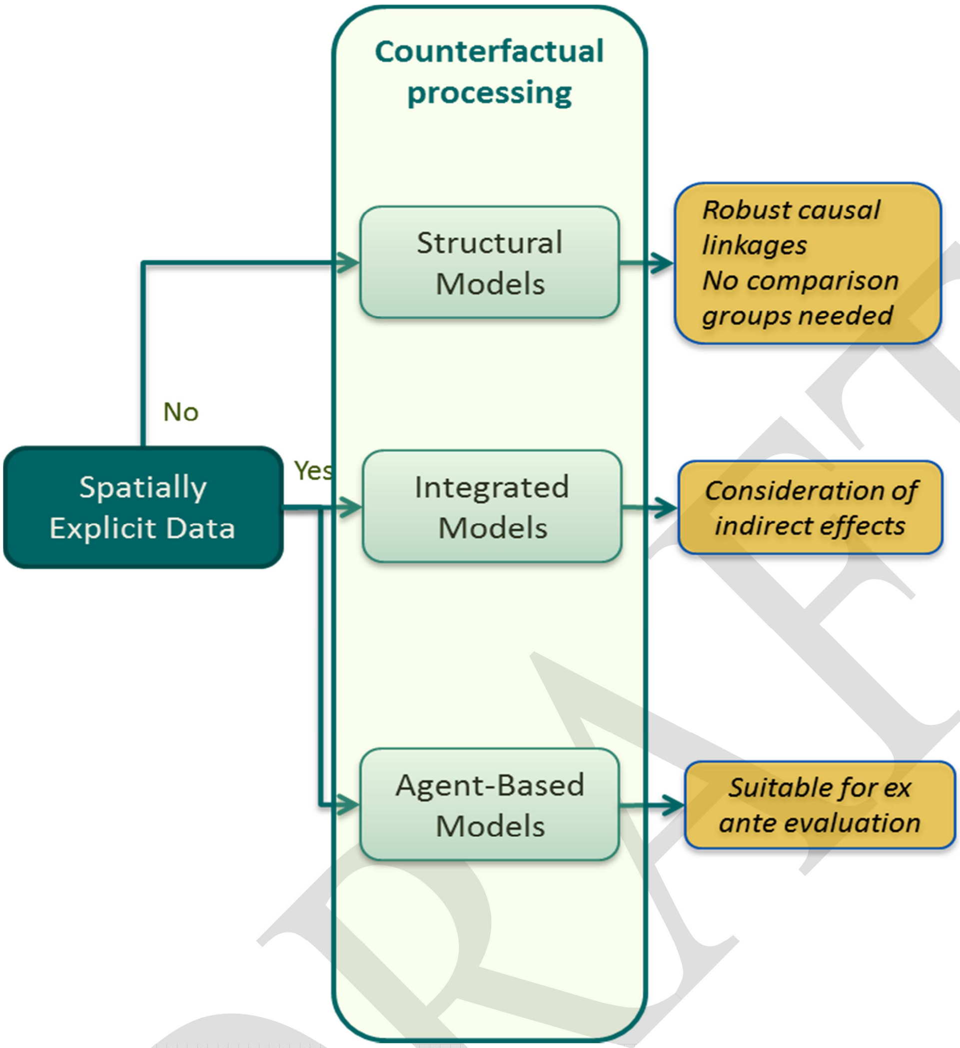

Figure 4.3: Long Run Evaluation Options w/o comparison groups (Step 3.3a - micro level logic model)

Step 3.3b - Naïve Estimates of Counterfactual

Naïve estimates of counterfactual should be used when data on programme beneficiaries prior and after programme are available, without to use particularly complex modelling approaches. It can be divided in two different techniques: (i) the naive "before-after" estimator, which utilizes programme data on programme beneficiaries to compute programme outcomes for programme participants (without counterfactual); and (ii) the naïve "with" vs. "without" approach, that use the non-participants as a control group.

These approaches are based on the assumption that in the absence of the programme, the outcome indicator of the participants at the programme would be the same as for non-participants at the same programme. The control group in the naïve comparison of programme beneficiaries is represented by the population average, including participants and non-participants. In this evaluation approach the data necessary for the on average outcome indicators in the group of non-participants is usually obtained from statistical databases. In this specific counterfactual design the possible methods linked with the naïve approaches at micro level are the ecological footprint, the integrated models and ABMs. In the first case no spatial explicit data are necessary. With the ecological footprint, and more in general with the use of composite sustainability indicators, it is possible to count the farm heterogeneity due to human environmental action to define better policy evaluations within a single agricultural system. In case of availability of spatial data integrated models and ABMs should be used, which have the characteristics described in the previous section. Basically the use of all these methods allows designing the counterfactual on the basis of the data commonly available from official statistical sources available at local level (e.g. FADN, Census, FSS).

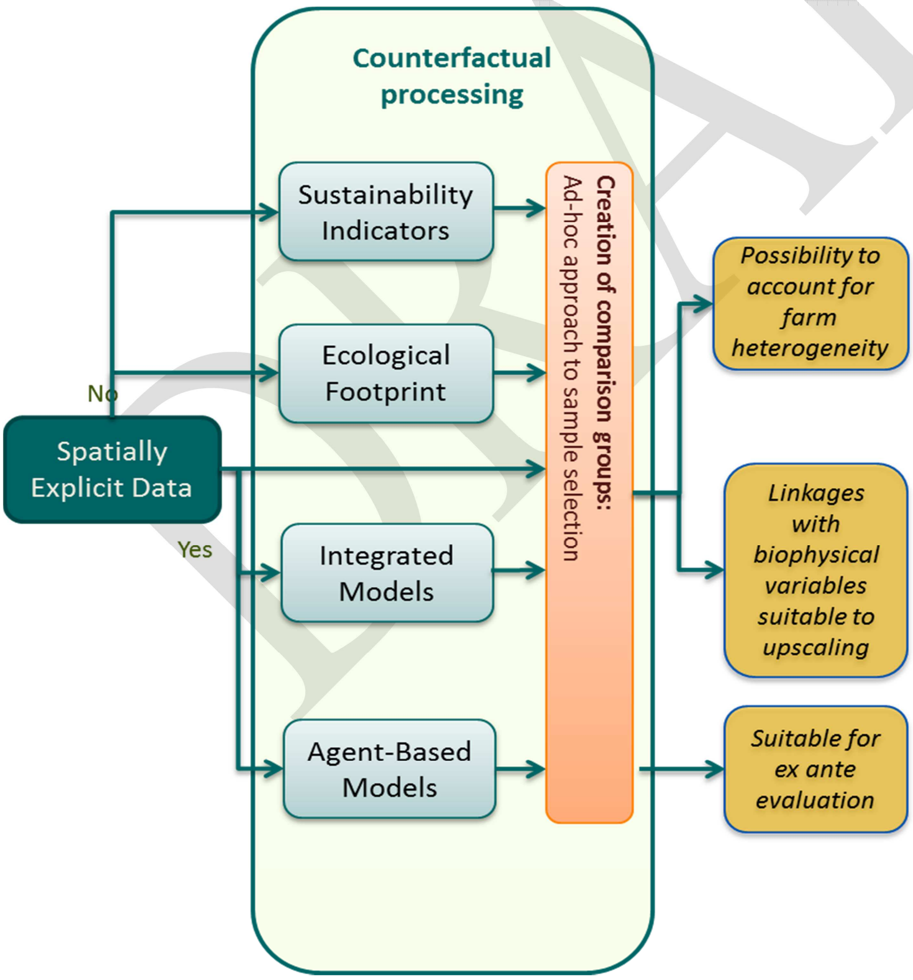

Figure 4.4: Naive Estimated of Counterfactual (Step 3.3b - micro level logic model)

Step 3.3c - Elaborate Statistics-based Evaluation Options

For this method is essential the abundant data availability about general characteristics and performance of beneficiaries and non-beneficiaries, before and after implementation of the RDP. The main techniques used to implement this approach are: the Difference in Differences (DID); the regression discontinuity design (RDD); the matching methods and propensity score matching (PSM); and the combined methods. The DID compares the before and after changes of programme participant and after change of outcome indicators. This approach allows to control the unobserved heterogeneity (under the assumption that this does not vary in time). This method requires data availability between two periods observed (time series). The RDD require availability of dataset with variable and observation on eligible and non-eligible units, with time series of cross-sectional data. In fact the RDD allows assessing the effects of programmes that have a continuous eligibility. The matching methods, including the PSM, are the most advanced and effective tools of evaluation. They are based on advanced statistical approaches and need of abundant data on participants and non-participants, requiring high quantitative skills of evaluator. Through this approach at the micro level, the methods selected are the ecological footprint and the integrated models. In the first case no spatial explicit data are necessary on the contrary in the integrated models explicit data are necessary.

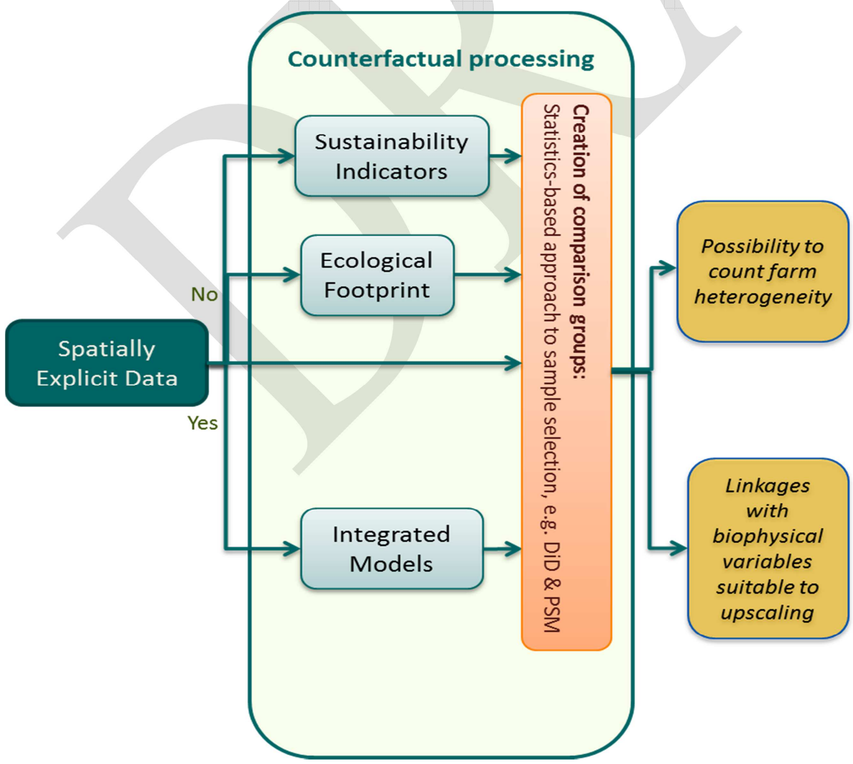

Figure 4.5: Elaborate Statistics-based Evaluation Options (Step 3.3c - micro level logic model)

Step 3.4 - Micro-Macro aggregation and validation

Multiple data sources are required for micro-macro level evaluation, deriving from different databases and providing for different data with different metrics and terminology. Regarding the terminology to be used, the farm can be defined as the baseline unit for micro-level analysis. However, it has to be underlined that the 'farm level' cannot correspond to the same meaning in different evaluation exercises. Macro and micro levels are used in different scales to identify different levels and therefore their use can be ambiguous. As highlighted before, in the evaluation assessment micro level is substantially represented by the farm, which is considered as the simplest management unit of the agricultural system linked to the implementation of RDP measures.

Each model can be more suitable for micro or macro level evaluation if a consistent aggregation procedure is available for the analysis. From a micro level point of view, spatial aggregation will consist of up-scaling and aggregating data from farm level to regional or national ones. However, micro-macro linkage can be difficult to detect, in relation to the criticisms, in ensuring the representativeness of assessed data to the universe of farms. Although up-scaling could facilitate the consistency in micro-macro linkage aggregation, it has to be highlighted that the occurring risk of summarising micro-level data for a macro-level perspective cannot always be ensured to represent the complexity of the universe of the agricultural systems.

Net impact evaluation at micro level can be ensured if indirect effects have been taken into account. In case of environmental impacts at micro level, deadweight effects seem to be more relevant if land use and practice changes would have even occurred without the intervention. Micro-macro linkages can lead to a better definition of indirect effects at macro level in presence of spatially explicit data in the case of environmental impact assessment.

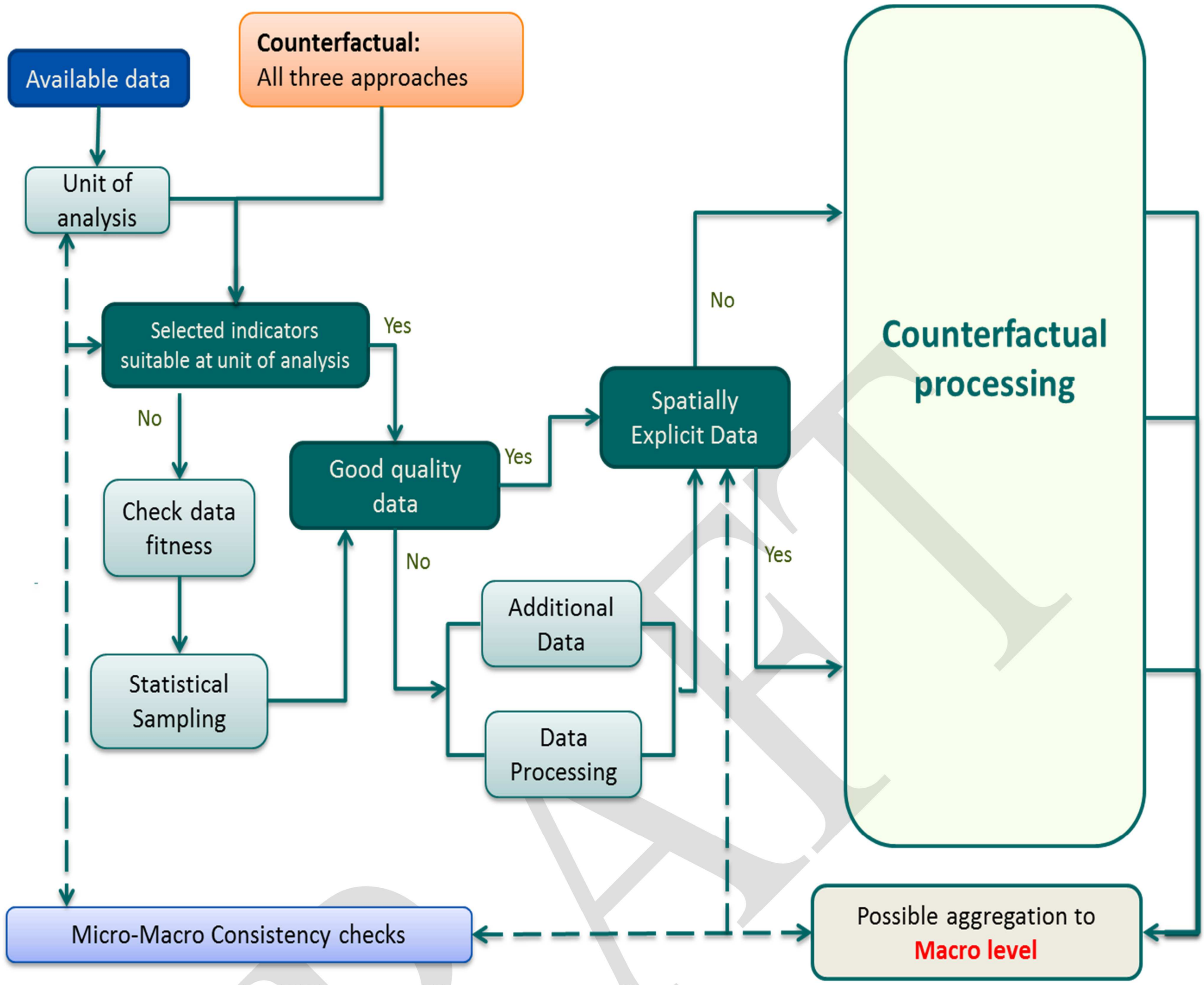

Figure 4.6: Micro-Macro aggregation and validation (Step 3.4 - micro level logic model)

5 Workflow and description of the macro level model

At the macro level the evaluators will assess the wider impact of the farm level implementation of RDP measures. The macro level model aims to allow for assessment of beyond the farm boundary impact (i.e. spatial aggregation of impact) as well as an assessment of the impact of multiple measures (i.e. RDP aggregation). Where possible the impact assessment is based on a single indicator however there are circumstances where the use of multiple indicators enhances the impact assessment. Examples for multiple indicators could be the application of a set of indicators included in the Landscape Character Assessment or several resource-based and problem-oriented indicators to assess animal welfare impacts. The macro level logic model is in many ways similar but that of the micro level logic model (section 4), in that there is a macro level logic model for each of the three counterfactual designs and that the initial phases of these models are the same. However, the main difference is that the workflow includes methods for both single and multiple indicators, which will lead to the net impact assessment. The use of a single indicator in this process can enhance the evidence of causality between RDP measure and impact, while the use of multiple indicators can specifically focus on assessments of multiple criteria (either multiple indicators for single measure or multiple indicators for multiple measures). The consistency of the net impact assessment is achieved through micro-macro consistency checks which are included at critical points in the workflow. The checks shall ensure that the results of the policy impacts identified at macro level (i.e. at catchment, regional or programme level including impacts on beneficiaries and non-beneficiaries) are consistent with the impacts identified for beneficiaries at micro level.

Step 4.1 Definition of functional unit and the consistency of the indicators.

At the start of the macro level workflow an appropriate functional unit is considered for each of the counterfactual designs based on the available data and where available aggregated micro level results. The unit of analysis is a spatial area beyond the farm (micro) level where the impact of RDP on public good can effectively be measured. At macro level it integrates the available data for a specific public good and the micro level results which at this stage is an important step to consistent indicators for the macro level evaluation (**Figure 5.1**). For those circumstances where aggregated micro-level results are included in the assessment the consistency between micro and macro level data is at this stage checked. The main purpose of this phase is to secure the link and consistency between micro and macro assessment and increase the opportunity to measure true causality.

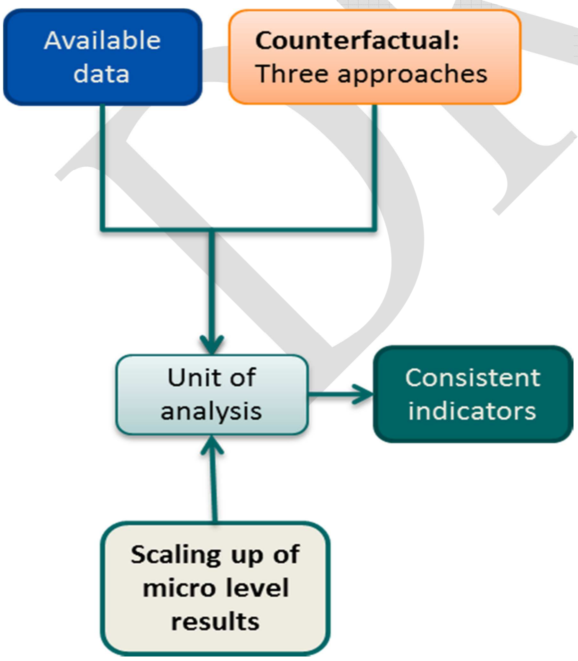

Figure 5.1: Definition of unit of analysis and the consistency of the indicators (Step 4.1 of macro level logic model)

Step 4.2 Creation of consistent spatial data

At this stage the macro level workflow leads to two possible paths towards the selection of evaluation methods based on the use of single or multiple consistent indicators, which can either be spatially explicit or not. If spatial data are available and used in the planned evaluation, these do not necessarily all have the same or an appropriate scale/resolution for the selected functional unit and method of analysis, where necessary scaling will precede the application of the selected methodology in the workflow.

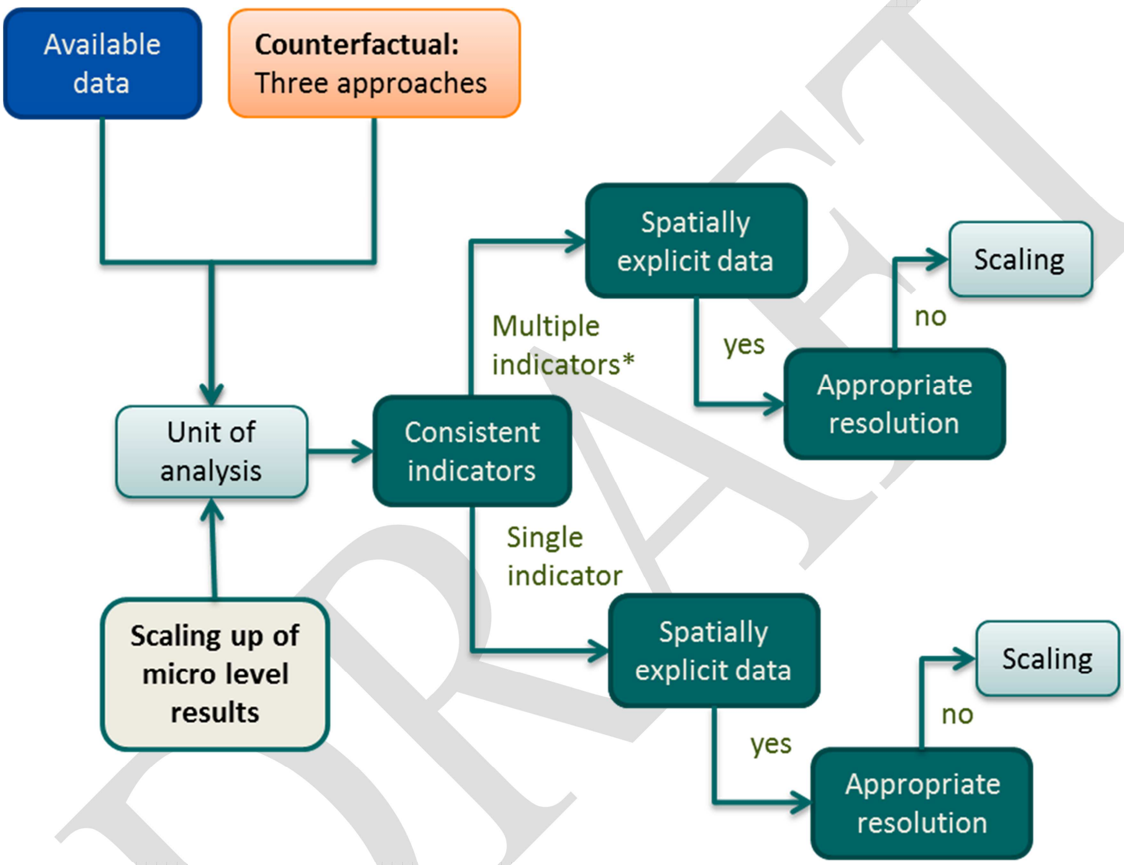

Figure 5.2: Step 4.2 Creation of consistent spatial data

Step 4.3a Long Run Evaluation Options w/o Comparison Groups

In the absence of a comparison group modelling methodologies will be used to improve our understanding of the macro level impacts. The selected modelling approaches can deal with single and multiple indicators **and internalize the counterfactual processes in their modelling framework**. The selected modelling methodologies contribute to the coverage of substitution effects, disentangling of external impacts and consistent micro-macro linkages.

CGE and PE modelling frameworks capture the policy impacts for the whole sector (PE models) or for the whole economy (CGE models) and can consider substitution effects within (PE models) and between different sectors (CGE models). CGE and PE modelling frameworks normally require larger scale applications and the availability of regional economic data sets to regionalise the modelling framework. Both modelling frameworks can be used with and without spatial data, whereby spatial modelling frameworks also enable the consideration of spatial substitution effects.

The availability of spatial data for single and multiple indicators would also allow the application of **spatial econometric models** (e.g. spatial lag models). Spatial econometrics enables to disentangle external impacts to assess the net-impact of the evaluated policy measures or programme.

Finally in the absence of comparison groups **integrated economic and biophysical modelling** can contribute to the understanding of causality links between impact of RDP on public good at both micro and macro level but in particular in relation to the consistency between micro and macro impacts.

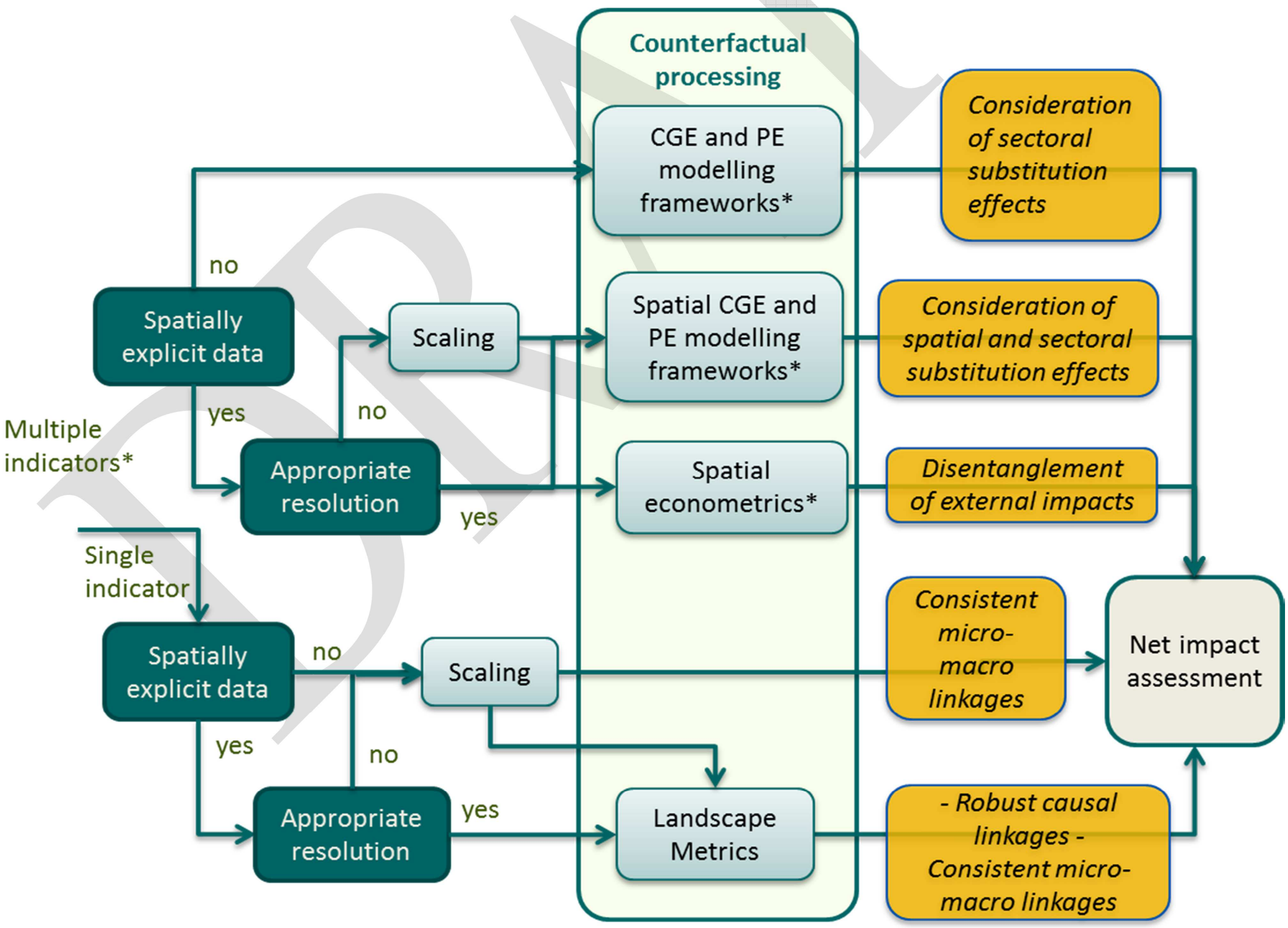

Figure 5.3: Step 4.3a Long Run Evaluation Options without comparison groups

Step 4.3b **Naïve Estimates of Counterfactual**

Naive estimates of counterfactuals are for those circumstances where there are data for the assesment but the comparison groups are not comparable or if they are comparable there is no timeseries and the variables explaing participation are unknown. The latter leaves an assessment based on with and without comparison.

For multiple indicators that are not spatially explicit two possible options are: **multi-criteria evaluation** and **ecological foot-printing.** These methods can be used to assess the macro-level heterogeneity. **Spatial multi-criteria evaluation** and **multifunctional zoning** are methods for an assessment with spatially explicit multiple indicators.

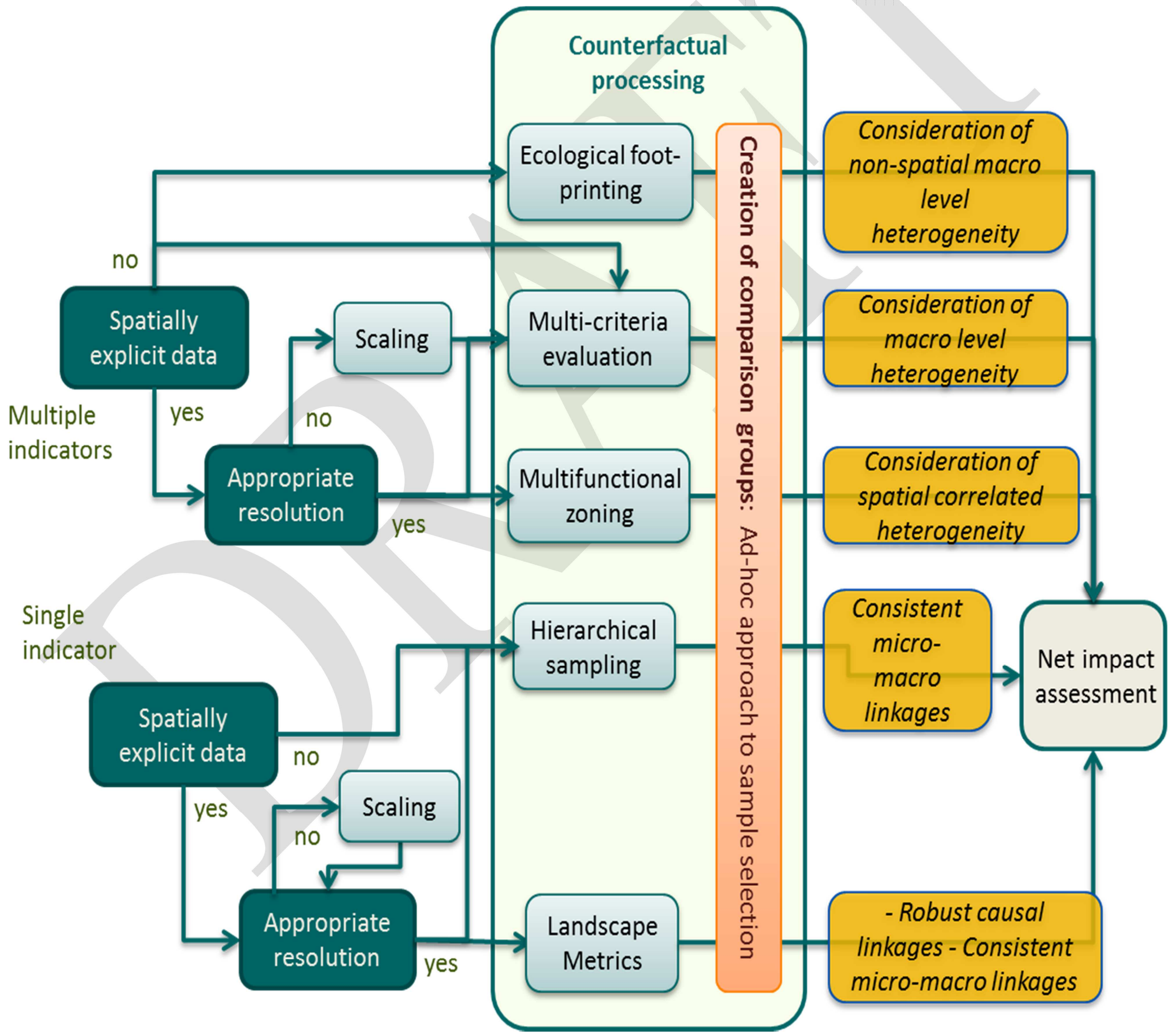

Figure 5.4: Step 4.3b Naïve Estimates of Counterfactual

Step 4.3c Elaborate Statistics-based Evaluation Options

In a data rich situation that allows for the construction of detailed comparison groups **elaborate statistics** can be considered for the evaluations.

For multiple indicators that are spatially explicit **spatial econometrics** is able to disentangle the actual impact of the policy measures from the impacts of different other drivers of change.

For single indicators there are methods for each of the non-spatial (**hierarchical sampling**) and spatial data (**spatial statistics/landscape metrics**), which support an improved assessment based on robust causal links between RDP measure and impact, as well as the consistency of indicators between micro-macro linkages. In addition based on the quality of the data these methods are able to consider non-linearity of impact (i.e. impact measured at micro level may not be measurable/effective or not in the same way at macro level).

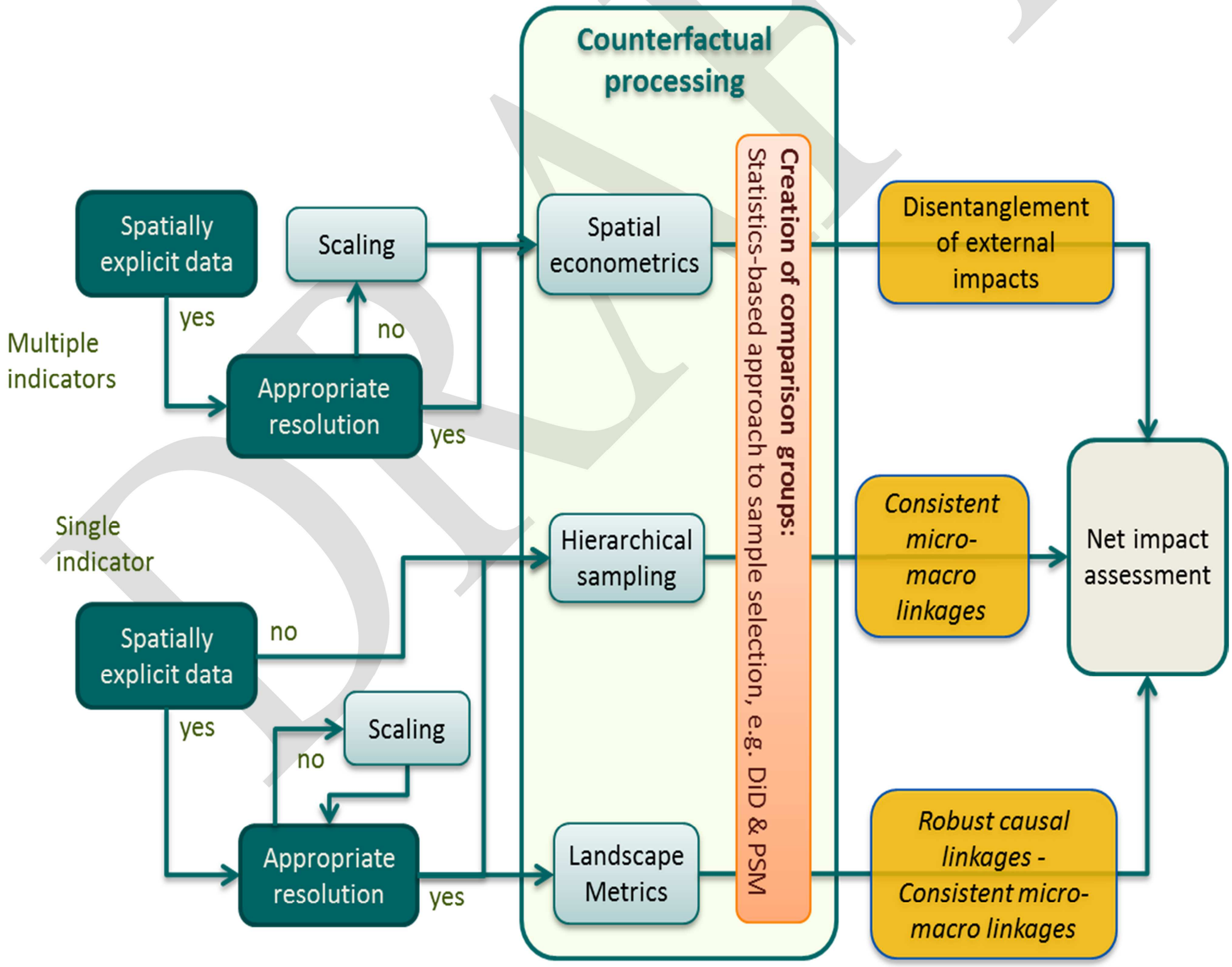

Figure 5.5: Step 4.3c Elaborate Statistics-based Evaluation Options

Step 4.4 Micro and macro consistency

The main purpose of the consistency checks at macro level are to ensure that the results of the policy impacts identified at catchment, regional or programme level (including impacts on non-beneficiaries) are consistent with the impacts identified for beneficiaries at micro level. The consistency checks depend on the type of macro level evaluations carried out. If micro level results form the basis of the macro level assessment, up-scaling and aggregating of the results and micro level data from farm level to catchment, regional or programme level could provide one option. Although up-scaling could facilitate the consistency in micro-macro linkage aggregation, it has to be highlighted that the occurring risk of summarising micro-level data for a macro-level perspective cannot always be ensured to represent the complexity of the universe of the agricultural systems. Indirect effects on non-beneficiaries such as substitution effects and displacement effects might not be taking into account adequately and the final assessment of net-impacts will rely on qualitative interpretation or (external) assumptions of indirect effects.

If the macro level assessment is based on regional / macro level data, consistency checks could vary (depending on available data) from qualitative interpretation of the consistency between the micro and macro level results to statistical comparisons and the use micro-macro consistency equations (e.g. in CGE models).

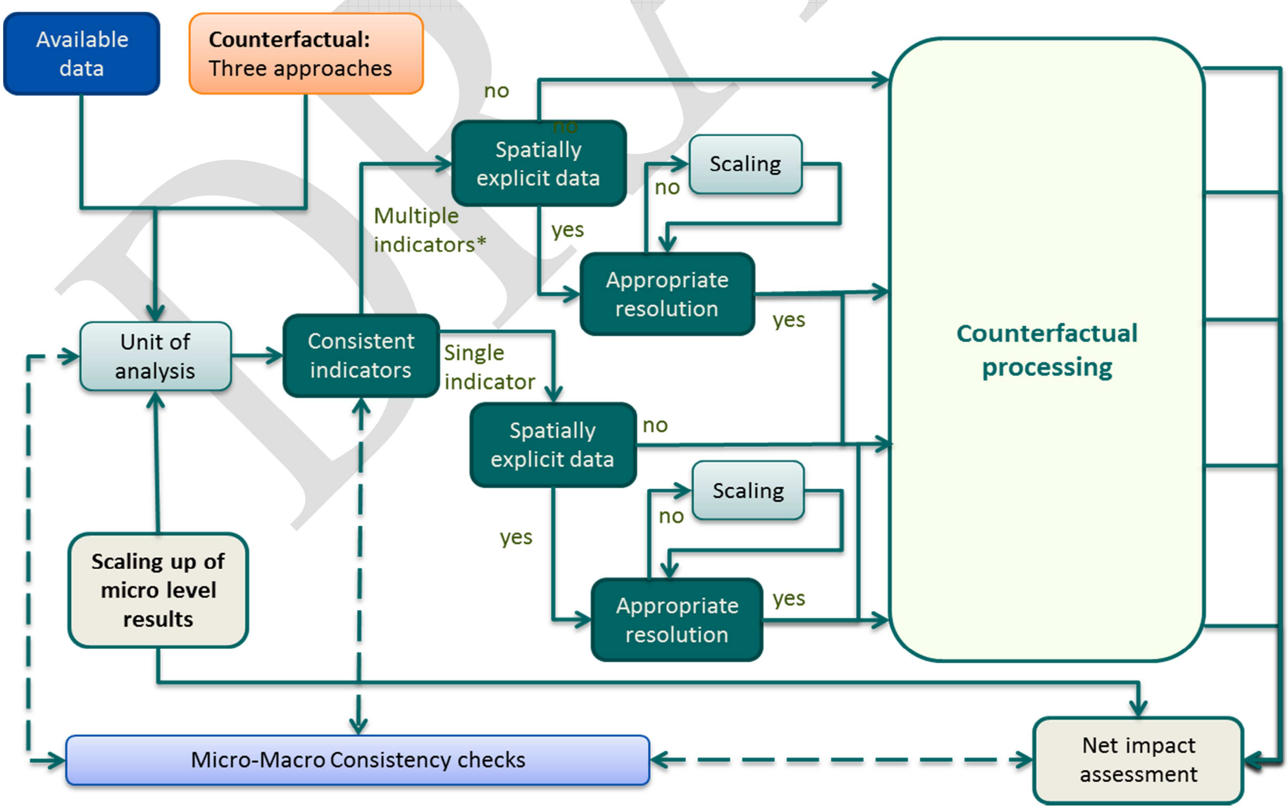

Figure 5.6: Step 4.4 Micro and macro consistency and net impact assessment

6 Concluding remarks

As reported in the previews sections the layers of the logic models explain the different steps in designing and carrying out counterfactual based micro level evaluations of environmental impacts of RDP. The evaluations are designed and linked with suitable indicators and methods on the basis of data availability and then options of consistency checks of data and results are reviewed. As an outcome of the logic model application, a consistent design of evaluation approaches can be achieved. This, however, is an iterative process and following the data collection the available data might not be sufficient for the initially selected evaluation approaches and revisions to the approach need to be carried out or a new method to be selected.

If suitable models produce sufficiently accurate results, we recommend their use. If a sufficiently accurate model does not exist or non-participants are in no way comparable to the participants (i.e. comparison groups cannot be separated in data), more qualitative methods are recommended or naïve quantitative methods have to be used in combination with qualitative methods. On the contrary, if there are time (data series) and resources (data availability and data quality) we recommend building a model or an advanced statistical approach that adequately deals with selection bias and enables a consistent counterfactual study of effects also in future evaluations.

Each step for the counterfactual brings a new method for the analysis. The subsequent steps check if existing models are available which can handle the case of no real comparison groups (i.e. cases where everyone in the evaluation area is a participant). The selection of an appropriate approach enables a self-updating system for evaluation with new data and possibilities to build on earlier modelling work.

Note that in most cases logic model appears to prefer quantitative approaches over qualitative. However, comparison group identification itself may be subject to qualitative assessment especially the more complex the setting gets, for example in the case of multiple comparison groups to assess synergies between measures or focus areas. Moreover, qualitative methods are essential to answer process-oriented questions in evaluations.

After the selection of an appropriate method following the logic model flowchart, the evaluator should know the type of model suitable to construct with the data at a general level. This consideration is then subjected to the micro or/and macro level specific logic models to further refine the evaluation possibilities. Please note, that even if the logic model would suggest a statistical approach, but the evaluator has access to a model, it is recommended that the cost-effectiveness of the different approaches is qualitatively assessed under the circumstances the evaluator is facing, or jointly implemented.

An important and complex aspect is the consideration of consistency checks between micro and macro level evaluations. If macro level results are generated through upscaling of micro level results, indirect effects on non-beneficiaries might not be taking into account adequately and the final assessment of net-impacts will rely on qualitative interpretation or (external) assumptions of indirect effects. If the macro level assessment is based on regional / macro level data, consistency checks could vary from qualitative interpretation of the consistency between the micro and macro level results to statistical comparisons and the use micro-macro consistency equations.