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

Report D3.2

Report on monitoring and data requirements for counterfactual methods

---

Author: Janne Artell (MTT)

Approved by Work Package Manager of WP3: Jyrki Aakkula (MTT)

Date: February 2014

Approved by Project Coordinator: Gerald Schwarz (TI)

Date: February 2014

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

List of Figures

List of Tables

List of Acronyms

BA	Before-and-after
DD	Double Difference
EMS	Economic Modelling & Simulation
FADN	Farm Accountancy Data Network
FSS	Farm structure Survey
GHG	GreenHouse Gas
GIS	Geographic Information System
IACS	Integrated Administration and Control System
IV	Instrumental Variables
PSM-DD	Propensity Score Matching Double Difference
RD-PM	Regression Discontinuity design and Pipeline Methods
RDP	Rural Development Programme
SEM	Structural Econometric Modelling
UAA	Utilised Agricultural Area
WW	With-and-without

Executive Summary

This report builds on the earlier review on counterfactual methods applicable in assessing the environmental impacts of rural development programmes and measures (Deliverable 3.1). This report elaborates on the types of data required for each method, the ways comparison groups can be formed to address major evaluation challenges, and provides an overview of the applicability of counterfactual methods to micro- and macro-level evaluation.

The report in particular elaborates on the different cases complicating simple with-and-without or before-and-after comparisons. The occurrence of participation status changes within the evaluation period, internal deadweight of earlier participation status and inertia of environmental effects, external deadweight of prior policies or environmental pressures each cause the number of comparison groups to increase. If such groups are significant in number, partial analysis of only participants and non-participants or other combinations will give biased impact estimates of unknown magnitude and direction.

Databases identifying factors affecting measure/programme participation probability are readily available for most statistical counterfactual analyses at micro level. Whether the existing data includes all the relevant factors affecting participation to specific measures is case specific (by measure, data quality and gaps), and may in some cases need to be appended by survey data. For macro-level analysis, micro-level data can be aggregated or low-level aggregation regional data used. Care must be taken to identify external drivers and pressures that affect participation status and environmental effects within and across regions.

The environmental indicators used in counterfactual analysis must relate causally, temporally and spatially as much as possible to the unit of analysis, e.g. farm or region. If environmental multiplier effects exist, they must be accounted for in the analysis. If that is not possible or the environmental impact indicator is very difficult to causally link to the actions at the unit of analysis, pressure indicators are recommended for use in evaluation. In some cases, pressure indicators can be transferred to environmental effects using biogeochemical modelling.

The complexity of the required high number of comparison groups to fully consider and assess net impacts emphasises the need for specific RDP monitoring programmes of environmental indicators (on participating and non-participating farms). While a larger number of comparison groups can generally be constructed based on existing secondary databases, data gaps on environmental indicators often constrain the use of complex counterfactual designs with a higher number of comparison groups in RDP evaluations.

1 Objectives of the Task

This report on data requirements for counterfactual methods builds on the earlier review on counterfactual methods applicable in assessing the environmental impacts of rural development at a programme and measure level. Data availability is the main driver for the applicability of different counterfactual methods. As such, this report elaborates on the types of data required for each method, the ways comparison groups[1] can be formed to address major evaluation challenges (e.g. long-term impacts and large scale implementation of policy measures, and imperfect data sources), and provides an overview of the applicability of counterfactual methods to micro- and macro-level evaluation. The identification of data requirements serves to show the applicable counterfactual methods in the WP6 case studies with the available data.

Figure 1 shows the linkages between tasks 3.2, 4.3 and 5.3 in connection with the choice of case study areas and the applicable methods. More specifically, the objectives of the comparison of the data requirements for counterfactual methods (Task 3.2) are to:

- inform the selection of case study areas in WP6 in terms of what kind of data need to be available in the areas to be able to test a method
- identify the key attributes for case study databases
- inform the development of the logic models and the selection of method combinations for the public-good case studies in the partner countries
- provide a list of methods for the selection of viable method combinations across WP3 – WP5 for the public good case studies.

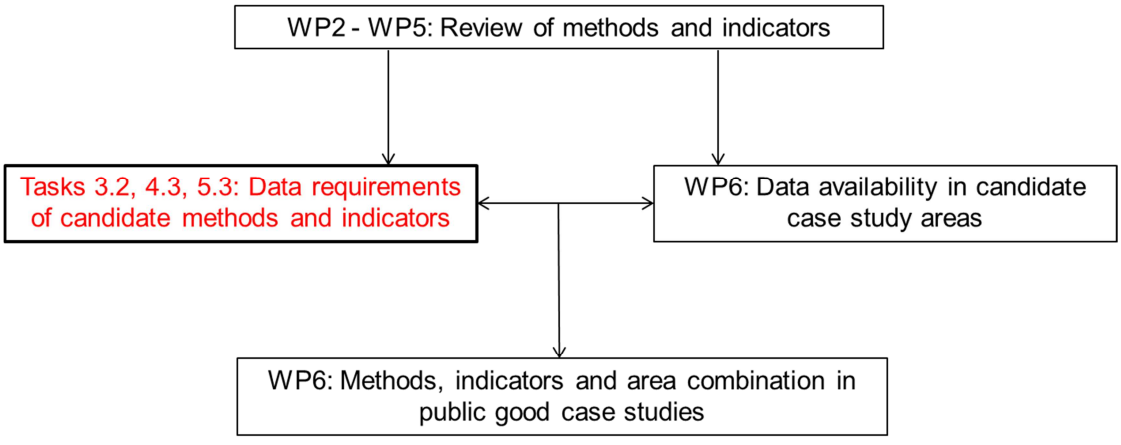

Figure 1 Overview of the different parts of the data assessment in the case study design

[1] The comparison group is often referred to as 'the control group'. However, the latter term suggests that the untreated (non-participant) group is randomly assigned. As this is rarely the case, we use the more general term 'comparison group'. In text we use 'participants' as shorthand for 'farms/regions that participate in the evaluated measure/programme' and 'non-participants' as the opposite term.

2 Assessment of the Data Requirements for Counterfactual Methods

Counterfactual methods can broadly be categorised into statistical approaches that typically consider farm-level impact evaluation, economic modelling approaches that can be tailor-made for micro- and macro-levels, and qualitative approaches that can be employed when quantitative data lacks in precision. Assessing the data requirements from the viewpoint of counterfactual analysis is not especially challenging; the analysis requires data on two or more comparable groups of farms or regions for which impact indicator data exists. Thus the key issues in counterfactual analysis lie in the applicability and sufficient availability of participant/non-participant data and indicator data.

In principle, any environmental pressure or impact indicator can be used in a counterfactual analysis as long as it has a causal link to farm-level or region-level measures. Problems arise when the assessed impact is not limited to the size unit of analysis (e.g. diffuse water pollution effects on water quality are very hard to attribute to farm-level data) and there exist multiplier or spatial spillover effects (e.g. biodiversity is not constrained to farm-level effects, and the effects may be strengthened by proximity to other participant farms). Using a non-specific indicator may under- or overestimate the environmental impact of the evaluated measure. The evaluator should thus choose an indicator that can best be related strictly to the unit of analysis. Sometimes this can be achieved by using pressure, rather than impact indicators. Impacts can then be assessed separately using environmental models, if such exist, that employ multiple stage transfer functions to transform pressure indicators (like fertiliser application) to impact indicators (water quality index). Multiplier effects may be, in some cases, covered by spatial econometric methods that can explicitly take into account effects between neighbouring participant/non-participant combinations, such analysis needs explicit data on neighbourhood participation status and the chosen environmental indicator either on the level of unit of analysis or at least on a buffer aggregate within a reasonable distance. Global impacts, such as climate change effects, are meaningless to assess at farm-level, thus making the use of pressure indicators like GHG emissions viable indicator candidates.

The statistical methods are subject to guidelines pertaining to the limits of all such approaches. A counterfactual approach requires that a comparison group exists, i.e. there are observations of farms or regions where the evaluated programme has not been implemented. Proper analysis requires a representative sample of farms/areas with enough observations of both groups for statistical analysis, the number of which increases by the complexity of the evaluated measure. Complexity is brought forward by sample selection issues. Sample selection is a problem when we compare participant and non-participant farms or regions and

assume that they are similar in all respects but the environmental indicator and participation. The statistical counterfactual methods try to correct for the biases appearing due to systematic (known) differences between participants and non-participants. These factors affecting the participation decision probability may be attributable to the unit of analysis (e.g. type of farm, farmer income and age, production method) or local conditions (e.g. prevalence of a biotope, soil type, micro weather).

Historical burden, or deadweight, can also affect both participation probability and the environmental indicator. External deadweight in the form of other directly or indirectly intervening factors such as other policies and other economic, social and environmental drivers should also be accounted for in the data if it may significantly affect participation and/or the environmental indicator status. Internal deadweight may occur when the same land management activities or investments would also be implemented without the policy measure and when the evaluated measure/programme has been ongoing before the beginning of the evaluation period. If there is reason to believe that prior participation to a measure/programme has influenced the decision to participate during the evaluation period and/or the environmental indicator value is dependent on the earlier participation status, there are at least four groups of comparison:

i) Earlier participants who currently participate
ii) Earlier participants who currently are not participants
iii) Earlier non-participants who currently participate
iv) Earlier non-participants who currently are not participants.

Thus, taking deadweight effects into account requires information on, at least, earlier participation status, the responsiveness and historical dependence of the chosen environmental indicator and significant outside factors affecting participation and causing environmental pressure.

Participation is not strictly restricted to the evaluation period. There may be drop-outs or late joiners to the measure/programme. In these cases the number of comparison groups increases. Specifying late joiners simply as participants and drop-outs as either participants or non-participants only biases the results of a counterfactual analysis. Provided that there are enough observations of such groups, they can be included in the analysis to provide further information on the effectiveness of the measure/programme when it has been partially implemented. Further, the measure or programme may change somehow during the evaluation period, affecting participation rates. In these cases, a proper counterfactual analysis requires

information of late joiners and drop-outs to provide an estimate of impacts for the whole evaluation period.

Finally, information on non-participants may be lacking due to various reasons (e.g. near full participation, data collection from only participants). Without a base for comparison a full-fledged counterfactual analysis is not possible using statistical approaches. In these cases three options may surface: 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) if there are queues to participation, comparing farms participating and those in queue together (pipeline method).

Table 1 summarises the minimum number of comparison groups needed in statistical analysis for assessing the full counterfactual effects with statistical methods. Note that in the case of all observations being participants, the counterfactual can only relate to a possibly existing subgroup. In these cases a full counterfactual analysis does not describe the question "*what would the word look like without the evaluated measure/programme*".

Table 1 Minimum number of comparison groups in different situations for full analysis

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 probability (x2)	None (x1)	6
			Historically significant outside pressure at min. one area (+2)	8
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

The additional number of groups of comparison provides chances to understand better the effects of the evaluated measure/programme. A full counterfactual analysis including all groups of comparison requires more extensive datasets on each group, but partial analysis can also be conducted by picking a suitable treatment group (e.g. full participants without internal or external deadweight) and comparison group (e.g. non-participants without internal or external deadweight). Partial analysis cannot be used to aggregate the impact analysis to the

whole population or would lead to an over- or underestimation of the net environmental impacts of the RDP. The likely consequences of evaluating data with generally known but unidentified special groups (due to data gaps) should be discussed after the analysis.

Farm-level ('micro level') data with information on participation status (and history) is often available in the hundreds and even thousands of observations, where problems arise in linking environmental outcomes to each observed farm. Regional data ('macro level') often includes environmental data at some comparable scale, but the number of comparable regions tends to be low for statistical analysis. Typically the statistical counterfactual methods identify participants completely, not partially, as is likely the case in regional analysis. If regions can be considered to participate in a programme fully or not at all and enough data exists for both groups to discern between inherent differences (size, special environmental conditions, earlier participation etc.) between the regions, the statistical counterfactual methods may be usable. However, the number of dimensions explaining the differences between regions also requires more observations from comparable regions. Depending on the size and hence the number of regions, a large number of observations may be impossible to acquire for analysis. If, on the other hand, most compared regions have partial uptake of the evaluated measure/programme, the counterfactual analysis requires a somewhat different approach. In such a case the question of the causal attribution of measure/programme participation to environmental outcomes in the region is vital to understand. The used regional indicators should not be affected significantly by unknown or undocumented effects to assess impacts in a meaningful way with statistical methods.

In economic modelling the existence of a comparison group may not be necessary if the model can be used to construct one. Actual data is used to calibrate the models to ensure realistic model predictions. In such cases environmental modelling is often needed, linking farmer behaviour on a regional or farm level to environmental outcomes.

Qualitative approaches are the most flexible, yet also the most multifaceted approaches. Data requirements for qualitative approaches pertain to getting the most relevant data available on the evaluated programme and subjecting this information to further scrutiny. Qualitative analysis can, however, be used in conjunction with other counterfactual methods to identify e.g. indirect links and interactions between policy measures and environmental outcomes.

Table 2 presents an overview of the type of counterfactual analysis and respective methods (with-and-without, WW, for ex-post analysis, and BA for before-and-after analysis of the evaluation period), special needs on participant/non-participant data for counterfactual

analysis, and the general applicability of the methods to micro- and macro-level evaluation. The 'disaggregated macro-level' applicability refers to spatial units with reasonably low aggregation of effects (e.g. municipalities). Each method is capable to cover multiple comparison groups (see Table 1). The suitability of indicators is case-specific and any indicator presentable in an ordinal, interval, or, in some cases, also categorical format can be used in analysis thus not affecting the interpretation of the table.

Table 2 Overview of the candidate methods, required data points in time and special requirements

Counterfactual approach	Method	Special requirements for participant/non-participant data	Applicability
WW	Propensity score matching, PSM (statistics)	factors affecting participation probability for each comparison group (farm income, type and size, earlier participation, farmer type, specific synergies from participation etc.)	Micro-level Disaggregated macro-level
BA	Double difference methods, DD (statistics)		Micro-level Disaggregated macro-level
WW & BA	Joint PSM-DD (statistics)	factors affecting participation probability for each comparison group (farm income, type and size, earlier participation, farmer type, specific synergies from participation etc.)	Micro-level Disaggregated macro-level
WW (& BA)	Instrumental variables regression, IV (statistics)	an instrumental variable that explains measure/programme participation probability for each comparison group but has no correlation with unobserved factors	Micro-level
WW (& BA)	Regression discontinuity design and pipeline methods, RD-PM (statistics)	measure/programme queue (on queue/out of queue) or eligibility information (eligible/non-eligible)	Micro-level
WW (& BA)	Structural econometric modelling, SEM (statistics /modelling)	depends on the model	Micro-level Macro-level
WW (& BA)	Economic modelling and simulation, EMS (modelling)	depends on the model	Micro-level Macro-level
WW (& BA)	Qualitative approaches, QA (partly non-numerical)	additional numerical analysis may require additional data, uses available data sources	Macro-level

Data sources for factors affecting participation, thus relevant for the development of comparison groups, would typically employ existing databases such as Farm Accountancy Data Network (FADN), Integrated Administration and Control System (IACS), Farm Structural Survey (FSS), or regional databases. Table 3 provides an overview of the main data

sources. Indicator data can be joined to the unit of analysis based on farm/region identifier information or GIS-based approaches. The complexity of the required high number of comparison groups to fully consider and assess net impacts emphasises the need for specific RDP monitoring programmes of environmental impact indicators (on participating and non-participating farms). While a larger number of comparison groups can generally be constructed based on existing secondary databases (see Table 3), data gaps on environmental impact indicators constrain the use of complex counterfactual designs with a higher number of comparison groups. Surveys can be used in some cases to cover gaps in data[2]. Reports D4.2 and D5.2 of the ENVIEVAL project elaborate on the different available indicators in micro and macro levels.

Table 3 Data requirements and sources of matching factors

Matching factors	Data format	Data origin	Temporal availability
Farm type (organic, crop, livestock, milk production etc.)	Database	FADN survey data FSS	annual case-specific decennial
Farm size (own / rented UAA, LU, production rates)	Database	FADN survey data FSS	annual case-specific decennial
Economic factors (farm income, full-time or part-time farming, labour, fertilizer/pesticide use, regional aggregates etc.)	Database Digital maps	FADN survey data FSS Eurostat	annual case-specific decennial semi-annual
Individual factors (age, education etc.)	Database	survey data FSS	case-specific decennial
Measure uptake	Database	IACS survey data	annual case-specific
Period of measure uptake (within and before current evaluation period)	Database	IACS survey data	annual case-specific
Biophysical factors affecting participation (location, topographic data, soil data etc.)	Database, digital maps, remote sensing and aerial photography	Multiple sources	
Intervening policies affecting participation (e.g. local incentives to participate)		Multiple sources	

3 Conclusions

This report has reviewed the data requirements for counterfactual methods. Importantly the report recognises cases complicating simple with-and-without or before-and-after comparisons. The occurrence of participation status changes within the evaluation period,

[2] Keeping in mind that specifically targeted surveys may lead to evaluator-generated sample selection bias in the counterfactual analysis.

internal deadweight of earlier participation status and inertia of environmental effects, external deadweight of prior policies or environmental pressures each cause the number of comparison groups to increase. If such groups are significant in number, partial analysis of only participants and non-participants or other combinations will give biased impact estimates of unknown magnitude and direction.

Databases identifying factors affecting measure/programme participation probability are readily available for most statistical counterfactual analyses at micro level. Whether the existing data includes all the relevant factors affecting participation to specific measures/programmes is case specific (by measure, data quality and gaps), and may in some cases need to be appended or replaced by survey data. For macro-level analysis, micro-level data can be aggregated or low-level aggregation regional data used. In macro analysis, care must be taken to identify external drivers and pressures that affect participation status and environmental effects within and across regions.

The environmental indicators used in counterfactual analysis must relate causally, temporally and spatially as much as possible to the unit of analysis, e.g. farm or region. If environmental multiplier effects exist, they must be accounted for in the analysis. If that is not possible or the environmental impact indicator is very difficult to causally link to the actions at the unit of analysis (e.g. from single farm actions to coastal water quality), pressure indicators are recommended for use in evaluation. In some cases, the pressure indicators can be transferred to environmental effects using biogeochemical modelling.