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 D3.1
Review of counterfactual methods

Authors: Janne Artell, Jyrki Aakkula, Heini Toikkanen (MTT)

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

Date: August 2013

Approved by Project Coordinator: Gerald Schwarz (TI)

Date: August 2013

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

AFP	Agrarinvestitionsförderprogramm
ATE	Average Treatment Effect
ATT	Average Treatment Effect on Treated
CGE	Computable General Equilibrium
CMEF	Common Monitoring and Evaluation Framework
FADN	Farm Accountancy Data Network
GIS	Geographic Information System
HNV	High Nature Value
IACS	Integrated Administration and Control System
ITT	Intent To Treat
IV	Instrumental Variable
PM	Pipeline Method
PSM	Propensity Score Matching
PSM-DD	Propensity Score Matching Double Difference
RD	Regression Discontinuity
RDP	Rural Development Programme
SWOT	Strengths Weaknesses Opportunities Threats
TBE	Theory Based Evaluation

1 Executive Summary

Impact assessment is an important tool for policy makers. By validating empirically the effects of a policy and revealing its strengths and weaknesses it enables better and more responsive policy design with intended effects. Impact assessment methodologies are well established in the literature existing both for ex-post and ex-ante impact evaluation. The environmental impact of a policy is the gap between the realised effects from policy implementation and a counterfactual scenario – state of the environment without the introduced policy.

This ENVIEVAL report reviews existing methodologies for counterfactual development, their use in previous and current RDP evaluations, and their applicability in ENVIEVAL WP6 public good case studies.

While there are multiple methods using counterfactuals for impact evaluation, there are challenges towards their use. Most methods require a decent amount of good quality data for proper analysis, which remains a key challenge to future evaluations. Data is needed on policy implementation effects on and by farmers and resulting environmental outcomes, measured by well-defined indicators. Further, the spatial variation of environmental outcomes needs to be taken into account explicitly: the benefits of many RDPs do not occur in the near vicinity of the participating farms.

A revision of national evaluations revealed that the description of counterfactual methods and impact indicators is generally vague. Counterfactual analysis has been conducted for the six public goods under inspection (climate, water quality, biodiversity: wildlife and HNV, soil, landscape). Water quality, wildlife and, slightly surprisingly, animal welfare are the public goods most often evaluated with a counterfactual approach, both within and across the countries under comparison, while measure 214, *Agri-environment payments*, is the single most often evaluated measure. With-and-without comparison of participants and non-participants is the typical reported approach, but there is little knowledge of the specific methods and control groups used to construct a counterfactual.

Based on literature reviews of methodologies and earlier evaluations, we recommend the use of well-established quantitative methods such as the Propensity Score Matching (PSM) and Propensity Score Matching Difference-in-Difference (PSM-DD) for environmental impact assessment. Both of these methods can overcome the biases suffered by naïve estimators given that sufficient information exists on the control group (i.e. non-participants) and time-related factors. It is important that baseline scenarios are created before the implementation of a programme to ensure an easier way to construct a counterfactual at the evaluation phase. In the case of ENVIEVAL WP6, ex-post public good evaluation case studies with existing baseline studies should be looked for, but may not exist.

The discussions at the stakeholder workshop and findings of the stakeholder consultation highlighted the need for more environmental monitoring data which consider the specific data requirements of RDP evaluations. In addition, the use of counterfactuals and the construction of control groups are often hindered by a lack of data for non-participants. Generally, a GIS-platform would provide a common base for statistical data storage easy to collate for further analysis. GIS-based data provide valuable information to impact evaluation methods as they can improve and reveal challenges in the formation of a counterfactual. Among other listed recommendations for the case studies to address, we recommend that relevant spatial information to each public good case be reviewed and added to the analysis.

2 Introduction

Impact assessment is an important tool for policy makers. By validating empirically the effects of a policy and revealing its strengths and weaknesses it enables better and more responsive policy design with intended effects (Lukesch & Schuch, 2010). Impact assessment methodologies are well established in the literature existing both for ex-post and ex-ante impact evaluation. In policy evaluation the weight is often on ex-post assessment. Constructing a counterfactual is a key element of ex-post environmental impact assessment, whereas ex-ante analysis explores impacts of different policy scenarios in the future versus a projected business-as-usual scenario (see figure 1). The environmental impact of a policy is the gap between the realised effects from policy implementation and a counterfactual scenario – state of the environment without the introduced policy.

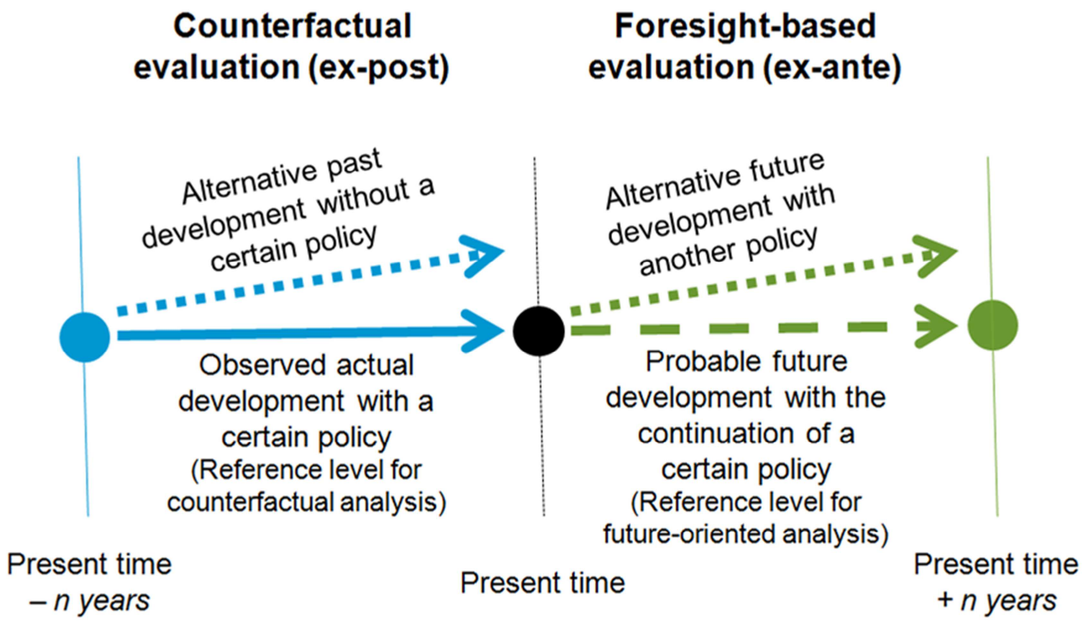

Figure 1 Ex-post and ex-ante impact evaluation

Counterfactual analysis includes many possible methodological approaches for quantifiable impact assessment, each method's applicability depending on the availability and quality of data. Counterfactual policy analysis should thus be given thought even prior to the policy implementation: measures or indicators of environmental impact must be known; monitoring the change in the chosen indicators and relevant attributes of the targeted and affected actors of the policy need to be recorded. The Common Monitoring and Evaluation Framework (CMEF), established during the most recent evaluation period 2007-2013, is the basis for the monitoring and evaluation of 2nd pillar agricultural support programs. The CMEF has a collection of common indicators split into four hierarchical categories: baseline, output, result and impact indicators. In addition, Member States are allowed to establish additional indicators which better describe national Rural Development Plan (RDP) targets. A functioning set of indicators should enable uniform impact assessment approaches using counterfactual analysis across Member States.

This ENVIEVAL report aims to:

- review existing methodologies for counterfactual development and their use in previous and current RDP evaluations
- review methodological developments addressing current challenges in counterfactual development and application
- recommend candidate methods for case studies on public good provision in ENVIEVAL WP6.

We first discuss the basic concept and challenges in forming a counterfactual. RDP evaluation literature[1] is subsequently reviewed to assess the current state of the art, level of reporting and challenges in using counterfactuals. We then provide an overview on the quantitative and qualitative methods used in counterfactual analysis based on the expert methodological introductions by Khandker et al. (2010) and Leeuw (n.d.). Finally we recommend methodological approaches for ENVIEVAL WP6 case studies on public good provision by RDPs.

3 Basic Concepts and Main Challenges in Counterfactual Development

3.1 Impact Analysis and Counterfactuals in Ex-post Assessment

Ex-post impact analysis is essentially the comparison of two states of the world, the current status after a treatment, e.g. an RDP measure, and a counterfactual status without the treatment. As the counterfactual is unobservable, it needs to be estimated using with real-world data or a theoretic framework. If the treatment, e.g. participation in an RDP measure, would be random in the study population, the environmental impact assessment for the programme would be relatively simple. The impact is shown in figure 2 as the distance between the environmental indicator values of the treatment (participants) group and the control group (non-participants), $Y_2 - Y_1$. The voluntary nature of RDP measures and

[1] Countries reviewed are Austria, Finland, France, Germany, Great Britain, Greece, Hungary, Italy, Lithuania and the Netherlands.

equality issues on measure allocation, however, prevent an experimental impact analysis based on perfect randomisation.

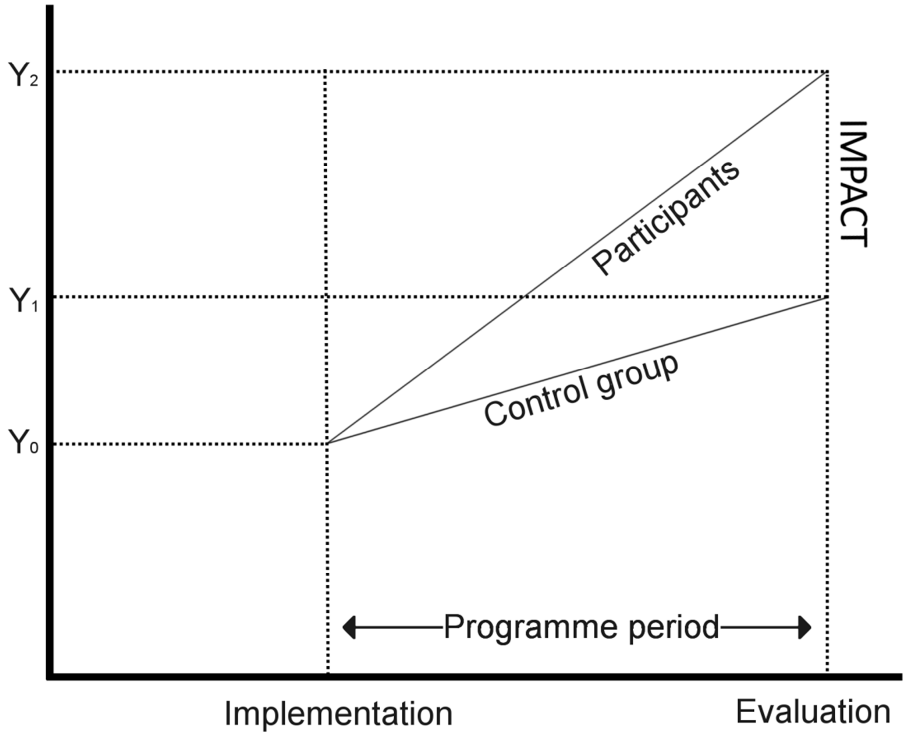

Figure 2 Environmental impact and a counterfactual (Adapted from Khandker et al., 2010)

3.2 Self Selection and Treatment Effects

The most important issue that a randomised experimental impact assessment avoids is sample selection. The data used to construct a counterfactual should represent the population under analysis as well as possible. Simple statistical analysis relies on the assumption of a random sample, meaning that there should not be any underlying dependencies on observed or unobserved factors in programme participation, i.e. self-selection to the programme (Khandker et al., 2010). Observed factors are, for example, rules limiting participation on the basis of farm type, size or location, while unobserved factors affecting programme self-selection could be anything from unique local conditions to cultural issues and mixes of regional policies. In an impact analysis comparing a treatment group and a control group from a different population, the assessment of an environmental impact will be biased up- or downwards depending on the situation. Figure 3 shows a case where the non-participant control group starts from an initially better environmental indicator level than programme participants before programme implementation. If the differences are not controlled for, the assessed environmental impact [$Y_2 - Z_1$] will be smaller than the unobserved actual effect [Y_2

– Y_1]. The size or direction of the sample selection bias is unknown without further information on the characteristics of actors.

Impact evaluation methods identify typically the expected values of these two impact types on the population level (Caliendo and Kopeinig, 2008, Khandker et al., 2010). The average treatment effect (ATE) is the difference between expected programme outcomes, *I*, of participants and non-participants and can be written as:

$$ATE = E[I_1 - I_0],$$

where subscripts 1 and 0 denote participants and non-participants, respectively. As discussed before, ATE does not represent the programme's impact correctly if self-selection into the programme occurs. The average treatment effect on the treated (ATT), on the other hand, measures the impact on the likely programme participants, i.e. excluding those who would never participate in the programme (Wooldridge, 2002). The ATT can be written as:

$$ATT = E[I_1|_{P=1} - I_0|_{P=1}]$$

i.e. the (unobserved) expected impact conditional on being a participant, $P=1$.

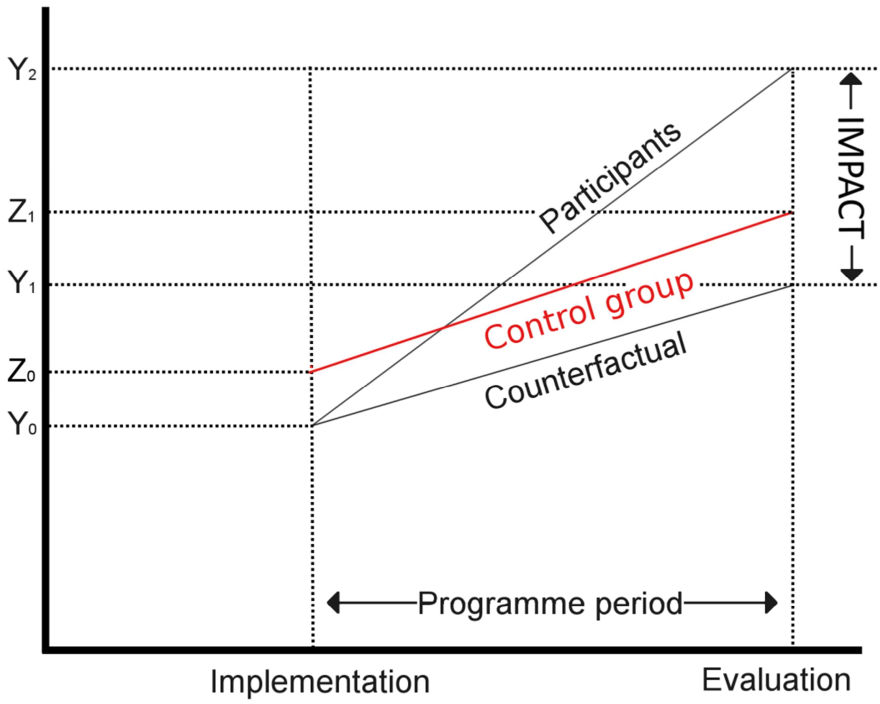

Figure 3 Programme self-selection bias (Adapted from Khandker et al., 2010)

3.3 Other Issues

Mandatory participation (e.g. the Nitrates Directive, cross-compliance) and very popular uptake of programmes (e.g. Agri-Environmental Measures in some Member States) represent cases where impact analysis using a control group is not possible. Then a comparison of the state of the world before the implementation of the programme and the status at the time of evaluation is tempting. A naïve counterfactual assumes static baseline environmental status over the programme period. As the assumption may be serious, Khandker et al. (2010) suggest a baseline study[2] on the participants to reduce biases in the counterfactual. The longer the estimated programme period, the more important it would be to take into account also other factors affecting the environmental indicator, such as spill-overs from other programmes or policies. Further, the nature of the environmental impact indicator used needs to be understood – natural processes may be slow to react and have stepwise regime changes in time periods that do not follow programme periods. Thus old sins and deeds affects today's perceived impacts, while current actions may take a while to emerge as impacts.

3.4 Main Challenges to Using Counterfactual Analysis in Evaluations

While there are multiple well-defined methods (see section 4) for ex-post impact evaluations using counterfactuals, there are challenges towards their use. Most methods require a decent amount of good quality data for proper analysis. Data is needed on policy implementation effects on and by farmers and resulting environmental outcomes, measured by well-defined indicators. Earlier literature has identified challenges encountered in current evaluation exercises, for example Anonymous, 2011; Cooper et al., 2006; Elsholz, 2008; Michalek, 2012a; Lukesch and Schuch, 2010; Stolze et al., 2000:

- There are too many indicators for national administrations to handle
- The focus is too much on indicators on outputs rather than outcomes
- Causality between measures and impacts increases in vagueness with multiple intervening factors (e.g. CAP and regional policies)
- Identification of impacts is complex due to different spatial scales of environmental outcomes and RDP implementation regions
- There are gaps in data and existing data suffers from errors
- Data may be difficult to obtain and use due to rigid data storage choices and other restrictions
- Evaluation capacity within Member States is lacking.

[2] The baseline study can also be used as a basis to model the counterfactual similar to the construction of a baseline, or business-as-usual scenario in ex-ante impact analysis.

Many of these problems were mentioned in stakeholder interviews. A predominant factor standing in the way of using counterfactual analysis was the lack of data. The lack of data was often reported due to restrictions on data use, preventing, for example, farm-level data from being used. A significant factor was related to the lack of true impact indicators and baseline studies or monitoring data to provide information for counterfactual analysis. Stakeholders also complained that there is no knowledge of the vast amount of data sources, and that there are problems and significant costs attributed to merging the data for impact analysis. Only in a few cases were there issues related to evaluation capacity in the member states, and many welcomed the idea to have a guidance manual to best practices to counterfactual analysis in conjunction with a common, fixed data source.

4 Current Applications of Counterfactual Analysis in RDP Evaluations

We review the reported use of counterfactuals in the recent RDP evaluations in Austria, Finland, France, Germany, Great Britain, Greece, Hungary, Italy, Lithuania and the Netherlands. More specifically, the focus and priority of the review was given to reports in the following order:

i) 2007-2013 Ex-post[3] evaluation results, 3 cases with a counterfactual reviewed,
ii) 2007-2013 Mid-term evaluation, 39 cases with a counterfactual reviewed,
iii) 2007-2013 Strategic Environmental Assessment, 0 cases with a counterfactual reviewed,
iv) 2000-2006 Ex-post evaluation reports, 26 cases with a counterfactual reviewed,
v) 2000-2006 Mid-term evaluation reports, 3 cases with a counterfactual reviewed,
vi) Other relevant RDP evaluation reports, 8 cases with a counterfactual reviewed.

We report the use of counterfactuals when it has been stated explicitly in the evaluations. Thus the accuracy and specificity of the original evaluation reporting drives our review results. Generally, the description of counterfactual methods and impact indicators is rather vague in national evaluations. Of the methods introduced in the next section of this report, none are explicitly mentioned in the evaluations[4], making the assessment of typical methods used difficult. Michalek (2012a) reports a similar assessment, stating that 75 per cent of RDP

[3] At the time of writing these reports represent more post-Mid-term evaluations, rather than fully fledged ex-post evaluations since the programme period is still ongoing.
[4] The only explicit indications of methodologies are given in Hungary (simple significance testing between groups) and France (disaggregate analysis).

Mid-Term Evaluations assess impacts without reference to counterfactual analysis, and naïve evaluation techniques reigned in the previous programme period impact assessments.

In table 1 we have listed the number of evaluations with reported counterfactual assessment by affected public good and country. In the evaluation reports of Finland, Greece and the Netherlands the use of counterfactuals is not mentioned, and is thus missing from the review.

Table 1 Number of impact evaluations with counterfactual analysis by public good categories

Country	**Climate**	**Water quality**	**Biodiversity: wildlife**	**Biodiversity: HNV**	**Soil**	**Landscape**	**Animal welfare**	**Total**
AT	-	-	-	-	-	-	1	1
FI	-	-	-	-	-	-	-	-
FR	-	2	2	2	2	-	5	13
DE	-	6	4	-	6	1	-	17
UK	2	1	1	1	-	-	-	5
EL	-	-	-	-	-	-	-	-
HU	-	-	1	-	-	-	-	1
IT	4	10	5	2	2	4	10	37
LT	-	1	1	1	1	1	-	5
NL	-	-	-	-	-	-	-	-
Total	6	20	14	6	11	6	16	79

Table 1 shows that counterfactual analysis has been conducted for all the relevant public goods. Italian evaluations have most often been described to have a counterfactual approach, and are the only ones covering all the public goods for one country. Water quality, wildlife and, slightly surprisingly, animal welfare are the public goods most often evaluated with a counterfactual approach, both within and across the countries under comparison.

Table 2 lists the counterfactual approaches reported in the reviewed evaluation reports[5]. The types of counterfactual analysis methods are classified broadly into four categories: with-and-

[5] Appendix A describes more specifically the evaluation documents using counterfactual approaches for each public good, including information on the evaluated measure, type of environmental impact indicator used, reported sources of data and the reported methodology and control group used to construct a counterfactual.

without, before-and-after, quantitative modelling and not specified. With-and-without comparison is the typical reported approach, but there is little knowledge of the specific methods used to construct a counterfactual. The most comprehensive counterfactual type (see section Double-difference methods (DD)) which combines with-and-without and before-and-after has not been reported. The general vagueness of reported methodologies and Michalek's (2012a) previous findings suggest that naïve counterfactual estimators are prevalent in the evaluation reports. Naïve control group formation was typically reported in evaluations on impacts to animal welfare and landscape, whereas quasi-experimental control groups, or non-random treatment assignment was reported often in other public good evaluations. Random experiments were less often observed, however, with the exception of water quality impact evaluations in Germany.

Table 2 Reported type of counterfactual approaches applied in reviewed RDP evaluation reports

	Type of counterfactual approach applied				
Public good	**With and without comparison**	**Before and after comparison**	**Quantitative modelling**	**Not specified**	**Total**
Climate	4	-	-	2	6
Water quality	13	-	4	3	120
Biodiversity: wildlife	9	-	-	5	14
Biodiversity: HNV	2	-	-	4	6
Soil	7	-	4	-	11
Landscape	6	-	-	-	6
Animal welfare	12	4	-	-	16
Total	53	4	8	14	79

The most often evaluated RD measure using a counterfactual approach is 214, *Agri-environment payments*, i.e. a *set* of sub-measures rather than single sub-measures, reducing problems in disentangling individual measure effects. The downside of this approach is the inability to assess a single measure's impact and efficiency as a part of the set of measures. In most evaluations a reference to CMEF indicators is given, but the actual indicator and its environmental impact are often not clearly reported. Thus a challenge remains to promote more rigorous execution of CMEF guidelines. Further, FADN and IACS data are the most often reported data sources for analysis, but the national database category may include FADN and IACS data but not be clearly stated in the evaluation report.

From the relatively small number of counterfactual applications and the lack of specificity in reporting in current evaluation literature it is evident that there are still major challenges for the RDP evaluation literature to provide transparent and more accurate impact assessments.

5 Main Methods Using Counterfactual Analysis

In the following we briefly introduce methods usable primarily in ex-post impact assessment including examples on RDP-related literature, and discuss the methodological requirements on available data for further studies. We also briefly explore ex-ante and qualitative impact analysis, their approaches on alternate, counterfactual world views, and data requirements.

5.1 Quantitative Impact Assessment Methods

Non-experimental evaluation methods make different assumptions on the type of sample selection bias, as self-selection, i.e. voluntary participation, to the programmes is allowed (Khandker et al., 2010, Pufahl & Weiss, 2009). The reader should refer to literature for a more elaborate discussion (e.g. Wooldridge, 2002, Khandker et al., 2010) and step-wise instructions on method application (Khandker et al., 2010; Michalek, 2012a on PSM-DD method).

5.1.1 Propensity Score Matching (PSM)

The PSM method constructs a counterfactual assuming that the statistical distributions of the basic characteristics of programme participants and non-participants overlap, i.e. they are not from two distinctly different populations. Programme participants are statistically matched[6] to characteristically similar non-participants by calculating a propensity score for each pair of observations. The average difference between the treated group and the matched control group represents the ATT of the evaluated programme. (Wooldridge, 2002; Caliendo & Kopeinig, 2008; Khandker et al., 2010).

As with any statistical method the availability and quality of data greatly affects the applicability of the method. Caliendo and Kopeinig (2008) note that successful PSM application requires understanding the form of sample-selection, i.e. the analysts should know and have information on all the characteristics affecting self-selection to a programme (for example the existence of traditional biotopes at the farm and the uptake of non-productive

[6] Multiple methods of matching exist in the literature, see Khandker et al. (2010).

investment support). Conversely, PSM is less useful in cases where it is likely that the programme participation decision depends on characteristics with little or no information available. Based on earlier literature[7], Khandker et al. (2010) recommend data for PSM studies to be taken from a single source including observations from programme participants, those potential for participation and non-participants to avoid measurement errors.

5.1.2 Double-difference methods (DD)

The double-difference method family, also named difference-in-difference or DiD in the literature, allows for unobserved characteristics affecting self-selection into the evaluated programme (Khandker et al., 2010). The degree of freedom allowed by the model comes at the cost of additional data requirements. More specifically, panel data on programme participants and non-participants is required from the time before and after its implementation. By assuming that the unobserved self-selection characteristics are stationary, the before-and-after comparison theoretically nets out the bias. In other words, the DD-method assumes that over the programme period the marginal change in the environmental indicator would be the same for both participants and non-participants without the programme. The marginal change in the environmental indicator for the non-participant group, which is available in the data, is then added to the initial status of the participant group to provide the counterfactual. Thus the DD-method assumes that one, the selection bias is additive, and two, the bias does not vary over the evaluation period (Khandker et al., 2010).

The assumptions may not hold if the programme is targeted or exogenous shocks have affected the participant and non-participant groups differently. To improve the DD method it is possible to combine it with PSM to exploit the advantages from both methodologies by estimating the treatment effect between the participants and matched non-participants (see e.g. Michalek, 2012a). However, this approach requires more extensive data on the population characteristics, both before and after the programme implementation. It is also possible to adjust the DD-method to exogenous shocks during the programme period if data on a more normal period of time on the same individuals is available (Khandker et al., 2010). To use a DD approach Khandker et al. (2010) advise the use of balanced panel data, i.e. data from the same individuals, from the time before and after the programme.

[7] Heckman J, Ichimura H, Todd P (1997) Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies* 64 (4): 605–54.
Heckman J, Ichimura H, Todd P (1998) Matching as an Econometric Evaluation Estimator. *Review of Economic Studies* 65(2): 261–94.
Ravallion M (1998) Evaluating Anti-Poverty Programs. In *Handbook of Development Economics*, vol. 4.. Schultz T, Strauss J (Eds.) North Holland. Amsterdam: 3787-3846.

Pufahl (2009) compared the PSM-DD approach to alternate parametric regression techniques finding no large difference in the model performance between the compared methods. However, static parametric models significantly overestimated impacts compared models taking time explicitly into account. Pufahl (2009) notes that the matching procedure is simple to understand, an important aspect when explaining the results to policy makers and the method can handle unbalanced numbers of participants compared to the number of non-participants where parametric methods may struggle. PSM and PSM-DD approaches can also handle multiple environmental indicators in the analysis, while separate models for each indicator are required for parametric approaches. This advantage obviously grows with the number of indicators specifying the environmental impact. Finally, in contrast to a simple parametric approach, the matching procedure allows to capture differences in environmental impacts due to different farm and regional characteristics.

Michalek (2012a) proposes the use of PSM-DD method to answer the EC Common Evaluation Questions with an example of using the method in Slovakia to assess the SAPARD programme impacts using FADN data and in Germany (Schleswig-Holstein) to assess the AFP programme impacts using national farm book-keeping data. Michalek (2012a) finds that using proper statistical methods provides evidence for discarding the use of biased naïve estimators. However, this puts pressure on the data availability and quality. As a concluding thought the author reminds that: "*While quantitative methods are advantageous for estimating and comparing net-impacts of various RD programmes they should be complemented with qualitative methods that are very helpful to answer questions: WHY? these effects occurred/not occurred in a given magnitude. A right combination of those both approaches appears therefore decisive for improving the quality of evaluation studies*".

Pufahl and Weiss (2009) assess the treatment effects of participating in agri-environment programmes in Germany using a combined PSM-DD approach. They use a privately owned book-keeping database (LAND-Data) of some 32,000 farms that includes information on area under cultivation, sales, labour, capital endowment, farm chemical purchases and participation in the agri-environment programme. The data does not allow distinguishing between the effects of single agro-environmental measures. Pufahl and Weiss (2009) assess the environmental impact through multiple proxies. They find the area under cultivation and grasslands to be higher at participant farms. Livestock density and expenditures on fertilisers and pesticides are, conversely, significantly lower with participation. The authors call for more specific farm-level data to assess to where the grasslands have expanded. This request reflects an important factor in impact assessment: where does the impact occur. As the used and available environmental indicators are contributors to rather than pure environmental impacts, spatial accumulation of effects and environmental impact models gain importance. Finally Pufahl and Weiss (2009) remind that the results of the PSM-DD approach is dependent on assumptions to which theory has little guidance, and the matching method often loses many observations from the analysis. Thus employing the method, as with any statistical approach, requires care and a large enough dataset.

Chabé-Ferret and Subervie (2013) also use PSM-DD approach to assess the impacts of five separate agro-environmental schemes (AES0201 and AES0205 promoting crop rotation, AES0301 cover crops, AES04 grass buffer strips, and AES21 organic farming) in France. Data for analysis including 400 to 3,000 participants depending on the scheme and 60,000 non-participants are compiled from multiple sources; the statistical services of the French Ministry of Agriculture, administrative agri-environmental scheme participation information and agricultural census data. The authors state that data compilation required multiple steps in the absence of concise, readily available data usable with the PSM-DD method. Environmental impacts of the agri-environmental schemes were assessed using specific outcome indicators for each scheme: crop cover area, length of buffer strips, main crop cultivation area and its share of total usable arable area, number of crop types cultivated, an index of crop evenness, and the land areas on and being converted to organic farming. Chabé-Ferret and Subervio (2013) find that the studied agro-environmental schemes in France have promoted environmentally friendly practices. The authors study also crossover effects, i.e. joint effects of participating in two or more AESs, finding them insignificant. In essence the result eases the analysis as impacts calculated for each AES need not take into account participation to the other four AESs. For future work the authors call specifically for a spatial analysis of treatment effects and resulting social benefits.

5.1.3 Instrumental variable methods (IV)

Instrumental variables estimation is an option for constructing a counterfactual if unobserved characteristics affect programme participation. The instrumental variable approach hinges on the existence of an applicable instrument – a variable uncorrelated with the unobserved characteristics, but correlated with the probability of participation to the programme (Khandker et al., 2010). With a good instrument the evaluation bias caused by unobserved characteristics decreases markedly. IV estimation can also be used in a panel setting, thus taking into account unobserved time-related effects between participants and non-participants.

Finding suitable instrumental variables in econometric analysis is often difficult. If the instrument correlates with individual characteristics, observed or unobserved, related to programme participation the counterfactual will be biased, just as it would with a more naïve estimation method. A similar effect occurs if the instrument has only low correlation with programme participation. The quality of the instrument in the estimation stage can be established with tests available in most econometric software. Primarily, though, the choice of instrument should have a logical explanation for the correlation with participation.

It should be noted that the IV method constructs the counterfactual to the likely participants of the programme instead of the actual participants as in the PSM and DD methods. This distinction is due to the instrumental variable, i.e. the proxy for participation, not predicting participation perfectly. Khandker et al. (2010) name this treatment effect as ITT – intent-to-treat effect and introduce also other treatment effects present in the literature. It still remains, though, that the evaluator should be very confident on the used instrument's applicability.

Khandker et al. (2010) note that the programme design and its implementation rules may provide exogenous instruments that correlate with participation highly but less so with unobserved characteristics. In RDP impact evaluation the set rules for participation may provide a source for good instruments. This approach is further discussed in the next section when we discuss the regression discontinuity method.

5.1.4 Regression Discontinuity design (RD)

Regression Discontinuity (RD) method exploits information on the rules of voluntary programme participation (Khandker et al., 2010). The rules provide a distinct cut-off point below which participation is not possible. Comparing participants, eligible and non-eligible non-participants in the same neighbourhood, i.e. very close to the cut-off rule above and below, provides a way to overcome unobserved factors. Greenstone and Gayer (2009) note that the frequent use of the RD method for impact analysis is largely due to the transparency of eligibility for treatment. For example in the RDP setting, spatial rules can be applied including farms close to the border that defines eligibility to LFA support. A non-spatial example is a case where programme eligibility is dependent on the number of livestock units.

The RD approach is vulnerable to bias if the administrative rules bend in practice or they change over the evaluation period (Khandker et al., 2010). Another issue in regression discontinuity approach is that the number of observations may be limited in the neighbourhood of the eligibility rule. As the method estimates impacts at the margin, or the cut-off point determined by the eligibility rule, the counterfactual applies to a subset of the whole population. In essence, the impacts estimated are strictly applicable at the neighbourhood of the rule, and may not be generalisable (see Khandker et al., 2010 and Lukesch & Schuch, 2010, for further discussion).

5.1.5 Pipeline Methods (PM)

The Pipeline Method (PM) can be used in conjunction to RD methods as a special case when programme participation is delayed due to, for example, budget constraints (Khandker et al., 2010). In the PM approach, information on potential participants to the evaluated programme are used to construct the counterfactual. Potential participants are identified as they enter the queue, the pipeline, to participate in the programme. This information is then applied in the RD framework to provide a more accurate counterfactual. This approach needs available detailed information on the programme applicants and the time of acceptance to the programme. If environmental monitoring data on the potential participants is readily available, the RD+PM method can be very suitable in RDP impact evaluation.

5.1.6 Structural approaches

The methods discussed thus far have been reduced form approaches, that is, we have estimated the treatment effect without assuming a structure in the causal chain. Reduced form econometric modelling approaches are not the only way to assess a counterfactual in impact analysis. A structural model builds on economic theory to create a framework of causality, which can then be tested using real-world data. The structured form can help to identify policy impacts when there are a number of policies affecting the assessed outcome (Khandker et al., 2010). The structural approach is related to qualitative methods that assess the mechanisms of the programme. Instead of a descriptive mechanism description, a structural approach takes a definitive, mathematical approach to cause-and-effect relations.

Nauges and Laukkanen (2011) provide an example. First, they construct an economic model explaining the rationale behind the technological adoption and then an empirical econometric model assessing the impacts from no-till farming. Nauges and Laukkanen (2011) use FADN data compiled with national weather and fixed-asset price index data for the analysis. They use a two-stage econometric model controlling for self-selection explaining the differences between the adopters of no-till and conventional farmers. To assess the environmental impacts Nauges and Laukkanen (2011) use transfer functions from earlier national literature concerning nutrient and herbicide run-off. The authors find that no-till farming decreases the run-off of both nutrients and herbicides and suggest that programmes endorsing no-till farming should be targeted for best effects.

5.1.7 Economic modelling and simulation

The structural approach to statistical/econometric counterfactual analysis is connected to economic modelling and simulation. Computable General Equilibrium (CGE) models, regional econometric and input-output models can be used to generate counterfactuals (Michalek, 2012b) not only for ex-post, but also for ex-ante impact evaluation. Especially in ex-ante analysis, economic models predicting impacts can be very helpful. The economic models are less dependent on issues such as sample selection, as the models assume rational economic actors. However, the models need to take into account all the factors affecting programme uptake probability, and, in the case of environmental impact assessment, the environmental impacts of a shift in farming practices as required by the evaluated RDP programme. Constructing a detailed model that would reflect reality may be difficult therefore as there are many factors affecting farmer decisions, farmers across Member States are heterogeneous and the environmental impacts are not always straightforward to assess. Computable General Equilibrium (CGE) models can be modified to tackle the economic impacts of new policies, but may be cumbersome.

Regina et al. (2009) provide an example of an ex-ante impact assessment of greenhouse gas mitigation measures using a CGE model for Finnish agriculture. Two scenarios up to year 2020 are compared: a baseline scenario with projections of agricultural product market; and another scenario where agricultural land area is restrained, the area of organic soils in production is reduced and the use for grassland increased, and biogas production is supported. The study finds that of all measures, those targeting the cultivation of organic soils are most effective in reducing emissions. Further, the model predicts that the crop production patterns would change, possibly affecting the success of other programmes in addition to other economic and societal reverberations.

5.1.8 Spatial analysis

The methods introduced before are often used without explicit reference to the spatial nature of policy impacts. The spatial dimension has multiple points of contact with counterfactual impact analysis. Lukesch and Schuch (2010) remind that *"the complexity and site specificity* *of potential environmental impacts of RD programmes, the identification of control groups and the establishment of a situation with and without the programme in place are very difficult. Moreover the lack of clear systemic borders of effects may lead to less reliable results in both the test and control groups”*. Factors explaining participation to the evaluated programme may not appear clear from basic farm statistics, but even a simple mapping of participants and non-participants in relation to environmental conditions, soil type, other programme implementation regions etc, may give significant clues and improve analysis. Further, GIS mapping of environmental impacts is important; it may be that the spatial distribution of participation reveals impacts that are centred or too widespread to have the intended effects. In the stakeholder interviews it was reported that a simple GIS approach revealed in Scotland that some measures were targeted at areas with little or not even possible desired effects. A number of studies suggest efficiency gains from improved spatial targeting of policies on non-market (public) goods (Van der Horst, 2007), and that there have been few, if any, case studies on agri-environmental schemes. Van der Horst (2007) continues that proper modelling programme participation requires detailed farm-level data and includes data on non-economic motives.

All of the methods introduced earlier can be amended with spatial information. A rigorous spatial analysis comes, however, with the additional burden of increased data requirements on areas with relatively few observations for statistical comparison of participants and non-participants[8]. Randomised data collection may overlook small special areas and very popular programmes may lack a control group. Combined environmental and economic modelling, while not being as dependent on many observations, requires more complex modelling approaches. For example, Fezzi and Bateman (2011) construct a structural spatially-explicit model on English and Welsh land use and production patterns. They find that, while some agricultural statistics are abundant and even mapped, some data may be less refined, which in turn requires the simplification of the model. The resulting model shows that policy impacts related to the Water Framework Directive are *highly spatially heterogeneous*, thus giving weight to taking the spatial dimension into account also in ex-post RDP evaluations.

Adding the spatial dimension to the econometric analysis imposes also a non-trivial data requirement; a particular challenge is to acquire enough observations that cover the spatial variation of environmental outcomes and local conditions. Environmental data is, fortunately,

[8] E.g. if spatial variation is included in a matching procedure (PSM and PSM-DD), the matching may prove difficult for very specific definitions of similar observations. However, as spatial data comes as an *addition* to other statistical data, it should not make evaluation more difficult, but rather enable additional options.

increasingly mapped and even a simple mapping of the data and results of a non-spatial impact analysis may reveal hidden patterns and causalities.

5.2 Qualitative Approaches to Impact Analysis

While impact analysis concentrates on methods that provide quantifiable ex-post and ex-ante impact evaluations, qualitative methods also exist. A qualitative assessment is unable to give tangible impact analysis as Khandker et al. (2010) note, but has the ability to create a logical view of the factors affecting the impacts. As Leeuw (n.d.) states, the systematic identification of the mechanisms, the links between cause and effect, of a programme and its outcomes opens up the 'black box' of impact evaluation enabling better understanding of the programme effects.

Theory Based Evaluation (TBE) is an example of a qualitative approach to impact evaluation. The theories are essentially the logic or reasoning of an intervention policy as understood by those who make, implement and are affected by the policy. Leeuw (2012) describes TBE to have both a conceptual and an empirical component. The conceptual component first develops a model of a programme that describes the underlying mechanisms. The mechanisms are not limited to statistics, but include also thought structures and logic frameworks. The empirical component then studies the causality of the programme and the outcomes. Counterfactual analysis is also possible when no statistical data is available. Leeuw (n.d.) notes that TBE can contribute to the formation of counterfactual for example using a counterfactual history approach, expert judgment and hypothetical 'what if' questions. Common to these approaches is the interviewing of as many experts as possible to provide the most likely counterfactual state of world. Leeuw (n.d.) also recommends the use of contribution analysis that extends the theory of change approach to TBE, discussed briefly below.

The theory of change method is based on the concept that policy evaluation is difficult mainly due to the ´poorly articulated´ arguments underlying policy choices (Leeuw, n.d.). The method requires that the evaluator is in close contact with the stakeholders from the beginning of the programme design, so that the articulation of each step leading to the intended outcome can be argued and identified. Five steps describe the use of the method. In the first step, the final desired outcome needs to be agreed upon, being mindful that the timeframe between the desired outcome and the programme period may differ. In the second step, the intermediate outcome after the programme period is agreed upon with respect to the ultimate goal. In steps three and four, the programme implementation methods and outcome targets are further divided into manageable parts with stakeholder discussion. In the final step, resources to achieve the goals with chosen implementation methods need to be assessed and adjusted. Impact evaluation with the theory of change method requires that the logical framework built earlier is accepted by the stakeholders to represent the actual state of the world and logic. With a clear definition of steps, monitoring methods and intended outcomes, the impacts can be evaluated also during the programme implementation (Leeuw, n.d.).

Contribution analysis extends the theory of change method. Instead of defining a strict causal relationship between the evaluated programme and environmental outcomes, the aim is to gather evidence that decreases uncertainties around the programme contribution to the observed change. A contribution story is central to the method. The story compiles links and assumptions of programme implementation and impacts. It also shows the evaluator the factors that are little known and, possibly, under scientific debate or subject to multiple theories, or programmes. The contribution story is iterated with the acquisition of new and existing empirical evidence. Empirical data is collected using *key informant interviews*, *focus groups and workshops and case studies* (Leeuw, n.d.).

5.3 Review of Methods

In table 3 we have reviewed the introduced methods listing the general idea of each method, the issues they address, and key challenges of implementation. The most preferred way of evaluation is a quantitative approach given that data exists. In cases where programme participants and non-participants are likely to have a similar development during the programme period, simpler methods can be used that only address sample selection issues. When this is not the case, panel data, i.e. data before and after the programme implementation, is required for constructing a proper counterfactual. Also, if a programme encompasses the whole population, methods using panel-data (i.e. DD and panel-IV analysis) are required. If data is non-existent, one must resort to qualitative methods, including stakeholder interviews and expert opinion surveys. A proper approach to impact evaluation would begin before the implementation of a programme with a qualitative assessment providing a basis for a more structured quantitative analysis after the implementation. This approach would also provide a basis for monitoring appropriate environmental outcomes at the chosen observation areas, and thus a more accurate analysis in the end.

Table 3 Methods for counterfactual analysis reviewed

Method	Approach	Issues addressed	Challenges
Propensity score matching, PSM	Matches participants to similar non-participants for statistical analysis.	Sample selection.	Data needed on all factors affecting programme participation: suffers from unobserved characteristics. Data required from both participants and non-participants.
Double difference methods, DD	Compares the net change of environmental outcome over time for both participants and non-participants.	Sample selection; unobserved characteristics.	Assumes similar development of all characteristics between participants and non-participants over time. Data required from both participants and non-participants before implementation and at the time of evaluation.
Combined PSM-DD	Matches participants to similar non-participants, compares environmental status before and after the programme.	Sample selection; unobserved characteristics; varying group effects over time.	Data required from both participants and non-participants before implementation and at the time of evaluation.
Instrumental variables regression, IV	Uses an instrument to explain participation to the programme, which is unrelated to the unobserved factors.	Unobserved characteristics.	Finding a suitable instrument and data for it is difficult and may not work universally across member states.
Regression discontinuity design RD	Compares participants and non-participants close to the borderline of an programme eligibility rule.	Unobserved characteristics.	Natural cases with discontinuities may be difficult to find in RDP programme evaluation. Generalising the results to participants and non-participants not close to the eligibility border may be non-trivial.
Pipeline methods, PM	Used as an addition to RD, employs information on potential participants	Unobserved factors; provides a more specific counterfactual than	Natural cases with discontinuities may be difficult to find in RDP programme evaluation.

	queuing to the programme.	RD.	Applicable only in cases where a queue to participation exists. Generalising the results to participants and non-participants not close to the eligibility border may be non-trivial.
Structural approaches	A theoretical framework for programme participation is formulated and tested empirically.	Explicit controls on other affecting policies; other issues depend on chosen statistical approach.	Constructing a theoretical framework from scratch requires time and expertise. Existing frameworks should be checked for applicability.
Economic modelling and simulation	Construction of a causal economic model including incentives to programme participation and environmental outcomes.	Explicit controls on all participation decision variables; varying group effects over time; enables ex-ante analysis.	A simple model may not represent reality adequately, while a working complex model is very time-intensive to construct. Needs validation with real world data.
Qualitative approaches	Number of approaches establishing causal relationships between programme participation and environmental outcomes.	Causal relationships between intended environmental impacts and programme participation on a general level.	Needs statistical analysis to provide exact results. For best applicability, should be conducted before programme implementation to provide a basis for evaluation.

6 Recommendations

Based on literature reviews of methodologies and earlier evaluations and stakeholder interviews, we recommend the use of well-established quantitative methods such as the PSM and PSM-DD for environmental impact assessment of RDPs similar to Lukesch and Schuch's (2010) recommendations. Both of these methods can overcome the biases suffered by naïve estimators given that sufficient information exists on the control group (i.e. non-participants) and time-related factors. A time-related factor is for example, what Lukesch and Schuch (2010) describe as a deadweight effect, the momentum of earlier decisions from both non-participants and participants affecting the environmental outcome during the programme period, also mentioned as a problem by some stakeholders. As such, it is important that baselines are created before the implementation of a programme to ensure an easier way to construct a counterfactual, already recommended in the CMEF Guidance notes (CMEF Guidance note B – Evaluation guidelines, p. 5) and sorely called for by stakeholders. Baseline estimations at regular intervals for true impact indicators would enable counterfactual analysis (DD) even in cases where most farms are participants and deadweight of earlier programmes carry over to the evaluation period, again a serious problem reported by stakeholders. For forming a baseline, the Guidance notes recommend either a qualitative SWOT analysis or other type of quantitative ex-ante analysis. In the case of ENVIEVAL WP6 ex-post public good evaluation case studies existing baseline studies should be looked for, but may not exist.

The spatial structure of farm-level programme adoption and the areas where the environmental impact occurs should also be given attention as also discussed by Van der Horst (2007) to provide better guidance on targeting future policy actions. The statistical methods for counterfactual analysis can be amended with spatial information. GIS-based data are valuable information to impact evaluation as they can improve and reveal challenges in the formation of counterfactuals. We thus recommend that relevant spatial information to each public good case be reviewed and added to the analysis.

We recommend ENVIEVAL WP6 public good case studies using a counterfactual analysis method to address issues related to:

i) chosen counterfactual method versus other methodological options
ii) innovative control group formation and related data sources
iii) identifying a baseline or if impacts vary over time for control group and related data sources (literature /own assessment)
iv) location and patterns of resulting environmental impacts in relation to the study area
v) location and patterns of participants vs. non-participants
vi) other intervening programmes or developments affecting programme uptake or environmental impacts in the study area
vii) generalisability of results (micro/macro/local/regional/national)

In most current evaluations, the data used comes from sources that should be available in Member States, e.g. FADN and IACS data. Environmental monitoring data, also in GIS format, is available in member countries to some extent, e.g. Natura 2000 network reports, Farmland Bird Index data, and reports related to the implementation of the Water Framework Directive and Marine Strategy Framework Directive. If CMEF evaluation questions could be linked to these data consistently for each type of public good, it would enable a more uniform and explicit way of assessing impacts. This, however, requires that suitable data sources are identified, causally linked to each other, and monitored frequently. The discussions at the stakeholder workshop and findings of the stakeholder consultation highlighted the need for more environmental monitoring data which consider the specific data requirements of RDP evaluations. In addition, the use of counterfactuals and the construction of control groups are often hindered by a lack of data for non-participants. Generally, a GIS-platform would provide a common base for statistical data storage easy to collate for further analysis. Many stakeholders found the current mass of data either difficult to grasp or, worse, difficult and expensive to merge not least due to restrictions in data use. Thus it would be very important to develop such a common platform for data.

Generally, impact analysis could also benefit from using qualitative approaches early on in policy/programme design. Programme design with clear definitions of intended environmental outcomes, usable indicators and monitoring data, and programme design should enable a platform for common analysis methodology, while also providing a transparent and logical framework for policy design.

7 References

Anonymous (2011) Monitoring and Evaluation for CAP post 2013 Stakeholder Conference, Background Document. Brussels, 20-21 September 2011. European Commission, Agriculture and Rural Development. Available online (accessed 30.07.2013): http://web.aimgroupinternational.com/2011/monitoring_evaluation_cap-post2013/files/Background_document.pdf

Caliendo M, Kopeinig S (2008) Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22(1): 31-72.

Chabé-Ferret S, Subervie J (2013) How much green for the buck? Estimating additional windfall effects of French agro-environmental schemes by DID-matching. *Journal of Environmental Economics and Management* 65: 12-27.

CMEF Guidance note B – Evaluation guidelines. Available online (accessed 30.07.2013): http://ec.europa.eu/agriculture/rurdev/eval/guidance/note_b_en.pdf

Cooper T, Baldock D, Rayment M, Kuhmonen T, Terluin I, Swales V, Poux X, Zakeossian D, Farmer M (2006) An Evaluation of the Less Favoured Area Measure in the 25 Member States of the European Union. A report prepared by the Institute for European Environmental Policy for DG Agriculture. November 2006

Elsholz R (2008) Assessment of Second Pillar programs at the regional level. In: Proceedings of 109th Seminar of European Association of Agricultural Economists, 20-21 November 2008, Viterbo, Italy. Available online (accessed 30.07.2013): http://ageconsearch.umn.edu/bitstream/44792/2/eaaepaper.pdf (accessed 25th July 2013)

Fezzi C, Bateman I (2011) Structural agricultural land use modelling for spatial agro-environmental policy analysis. *American Journal of Agricultural Economics* 93(4): 1168-1188.

Greenstone M, Gayer T (2009) Quasi-experimental and experimental approaches to environmental economics. *Journal of Environmental Economics and Management* 57: 21-44.

Khandker SR, Koolwal GB, Samad H (2010) Handbook on Impact Evaluation: Quantitative Methods and Practice. World Bank. pp.280. Available online (accessed 30.07.2013): https://openknowledge.worldbank.org/bitstream/handle/10986/2693/520990PUB0EPI1101Official0Use0Only1.pdf?sequence=1

LAND-Data Ltd.. Bookkeeping data of agricultural enterprises in Germany. Available online (accessed 30.07.2013): http://www.landdata.de

Leeuw FL (2012) Linking theory-based evaluation and contribution analysis: Three problems and a few solutions. *Evaluation* 18(3): 348-363. Available online (accessed 30.07.2013): http://evi.sagepub.com/content/18/3/348.full.pdf+html

Leeuw FL (no date) Theory-based evaluation. Based on material produced for DG Regional Policy by Frans L. Leeuw. Available online (accessed 30.07.2013): http://ec.europa.eu/regional_policy/information/evaluations/pdf/impact/theory_impact_guidance.pdf

Lukesch R, Schuch B (2010) Approaches for assessing the impacts of the Rural Development Programmes in the context of multiple intervening factors. Working paper. Findings of a Thematic Working Group established and coordinated by The European Evaluation Network for Rural Development.

Michalek J (2012a) Counterfactual impact evaluation of EU rural development programmes – Propensity Score Matching methodology applied to selected EU Member States. Volume 1: A micro-level approach. European Commission, Joint Research Centre – Institute for Prospective Technological Studies, EUR number 25421 EN.

Michalek J (2012b) 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, Joint Research Centre – Institute for Prospective Technological Studies, EUR number 25419 EN.

Nauges C, Laukkanen M (2011) Environmental and production cost impacts of no-till in Finland: Estimates from observed behavior. *Land Economics* 87(3): 508-527.

Pufahl A, Weiss CR (2009) Evaluating the effects of farm programmes: results from propensity score matching. *European Review of Agricultural Economics* 36(1): 79-101.

Pufahl A (2009) Empirische Wirkungsanalyse direkter Transferzahlungen - am Beispiel von Agrarumweltmaßnahmen und der Ausgleichszulage für benachteiligte Gebiete. University of Göttingen, Dissertation. Available online (accessed 30.07.2013): http://hdl.handle.net/11858/00-1735-0000-0006-B043-1

Regina K, Lehtonen H, Nousiainen J, Esala M (2009) Modelled impacts of mitigation measures on greenhouse gas emissions from Finnish agriculture up to 2020. *Agricultural and Food Science* 18: 477-493.

Stolze M, Piorr A, Häring AM, Dabbert S (2000) Environmental impacts of organic farming in Europe. *Organic Farming in Europe: Economics and Policy* 6. Universität Hohenheim, Stuttgart-Hohenheim.

Van der Horst D. (2007) Assessing the efficiency gains of improved spatial targeting of policy interventions; the example of an agri-environmental scheme. *Journal of Environmental Management* 85: 1076-1087.

Wooldridge JM (2002) *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, Massachusetts.

8 Appendix A

In the following tables 4 through 10 we list, by public good, evaluation reports using a counterfactual approach including information on the evaluated programme or measure, reported impact indicator, type of counterfactual analysis, and data sources. The types of counterfactual analysis methods are classified broadly into **four categories**: unknown, with-and-without, before-and-after, and combined with-and-without and before-and-after approaches. Control group formation in the counterfactual analysis are also reported in **four categories**: unknown, naïve (no plan of data sampling), quasi-experimental (non-random treatment assignment), and experimental (random treatment assignment). Data source categories are divided into **six categories**: unknown, FADN, IACS, case study (includes surveys), prior literature, and national databases (and a possible descriptor).

Table 4 Counterfactual analysis in climate related impact assessment

Country	Report	Programme or measure	Impact indicator	Type of CF analysis	Control group formation	Source of data for the CF
UK, England	MTE 2007-2013	214	Indirect result and output indicators	Unknown	Unknown	Prior literature
UK, England	MTE 2007-2013	221	Indirect result and output indicators, some reference to CMEF	Unknown	Unknown	Prior literature
IT, Veneto	EXP 2000-2006	214 H / I Sub I (9.1)	CMEF, average annual net carbon storage from 2000-2012	WW	Quasi-experimental	FADN
IT, Veneto	EXP 2000-2006	214 H / I	CMEF, trend in average annual net carbon storage beyond 2012	WW	Quasi-experimental	FADN
IT, Veneto	EXP 2000-2006	214 H / I	CMEF, net carbon storage with fossil origin, storage between the 2000 - 2012	WW	Quasi-experimental	National database (Private: AIEL)
IT, Veneto	MTE 2007-2013	221	Reduction of CO2 emissions equivalent	WW	Quasi-experimental	FADN

MTE = Mid-term evaluation report | EXP = Ex-post evaluation report | EXA = Ex-ante evaluation report
WW = With-and-without comparison

Table 5 Counterfactual analysis in water quality related impact assessment

Country	Report	Measure or Programme	Impact indicator	Type of CF analysis	Control group formation	Source of data for the CF
DE, Rheinland-Pfalz	MTE 2007-2013	214, environment-friendly agriculture and organic farming"	CMEF, change of nutrient balance (GNB) and pesticide use	WW	Naïve	National database (GNB)
DE, Baden-Württemberg	MTE 2007-2013	214, extensive grassland management	Animal stock density, LU/ha	WW	Experimental	IACS
DE, Baden-Württemberg	MTE 2007-2013	214 N-B1: extensive grassland management	Animal stock density, LU/ha	WW	Experimental	IACS
DE, Baden-Württemberg	MTE 2007-2013	214 N-B2: extensive management of permanent grassland	Animal stock density, LU/ha	WW	Experimental	IACS
DE, Baden-Württemberg	MTE 2007-2013	214 N-B3: grassland management in mountainous areas	Animal stock density, LU/ha	WW	Experimental	IACS
DE, Baden-Württemberg	MTE 2007-2013	214 N-D2: organic agriculture	Animal stock density, LU/ha	WW	Experimental	IACS
UK, England	MTE 2007-2013	214	CMEF, surplus nutrients per ha, pesticide application, and result (6b) and output indicators (indicators 34, 35, 36 & 37)	Unknown	Unknown	Prior literature
IT, Veneto	EXP 2000-2006	F 2,3, Integrated farming, Organic agriculture	CMEF, unirrigated area percentage	WW	Naïve	FADN
IT, Veneto	EXP 2000-2006	F 9, Set aside	CMEF, unirrigated area percentage	WW	Naïve	FADN
IT, Veneto	EXA 2007-2013	214	CMEF, change of nutrient balance (GNB)	WW	Quasi-experimental	FADN
IT, Veneto	EXA 2007-2013	214	Nitrogen loading measure	WW	Quasi-experimental	FADN
IT, Veneto	EXA 2007-2013	214	Nitrogen loading measure	WW	Quasi-experimental	FADN
IT, Veneto	EXA 2007-2013	214	Phosphorus loading measure	WW	Quasi-experimental	FADN
IT, Veneto	EXA 2007-2013	214	Phosphorus loading measure	WW	Quasi-experimental	FADN

IT, Veneto	MTE 2007-2013	214 B, C, E	CMEF, reduction of nitrogen and phosphorus surplus in the areas of intervention	WW	Quasi-experimental	FADN
IT, Veneto	MTE 2007-2013	214 C	CMEF, reduction of "risk index" resulting from pesticide application	WW	Quasi-experimental	National database
IT, Veneto	MTE 2007-2013	214 A	CMEF, wooded buffer strips	Unknown	Unknown	National database
LT	EXP 2004-2006	212 and 214[9]	Water pollution	WW	Naïve	IACS
FR	MTE 2000-2006	Gestion des ressources en eau	Proportion of arable land with less than 170 kg/ha/y nitrogen fertiliser	Unknown	Naïve	National database (PDRN, RICA monitoring)
FR	MTE 2000-2006	Gestion des ressources en eau	Proportion of arable land under organic farming, integrated production, and pasture with less than 2 LU / ha	Unknown	Naïve	National database (PDRN, RICA monitoring)
MTE = Mid-term evaluation report | EXP = Ex-post evaluation report | EXA = Ex-ante evaluation report WW = With-and-without comparison						

Table 6 Counterfactual analysis in wildlife related impact assessment

Country	Report	Measure or Programme	Impact indicator	Type of CF analysis	Control group formation	Source of data for the CF
DE, Brandenburg	EXP 2007-2013	214-A3	Breeding success of meadow birds	WW	Quasi-experimental	Case study
DE, Brandenburg	EXP 2007-2013	214	Indicator plant species	WW	Quasi-experimental	National database (habitat directive monitoring), IACS
DE, Baden-Württemberg	MTE 2007 - 2013	214	Agricultural and forest area under support, agricultural and forest area under management contributing to biodiversity and HNV	WW	Quasi-experimental	IACS
DE, Lower Saxony	EXP 2007 - 2013	121	Change in grassland area	WW	Quasi-experimental	IACS
HU	EXP 2000-2006	214	Presence of common bird	WW	Quasi-experimental	National database

[9] Same counterfactual analysis was used for both measures.

			species			(BirdLife Hungary monitoring network)
UK, England	MTE 2007-2013	214	CMEF, change in trend in biodiversity decline, measured by farmland species population	Unknown	Unknown	Prior literature (Natural England report, scientific literature)
IT, Veneto	EXP 2000-2006	214 F 2, 3, 5, 9, 11, 12	CMEF, VI.2.A-1.3. evidence of a positive relationship between assisted input reduction measures on the targeted land and species diversity (description, where practical involving estimates of species abundance)	Unknown	Quasi-experimental	National database (Natura 2000 Farmland bird index)
IT, Veneto	EXP 2000-2006	214 F 8	CMEF, VI.2.B-2.1. assisted ecological infrastructure with habitat function or non-farmed patches of land linked to agriculture (hectares and/or kilometres and/or number of sites/agreements) (d) of which enhancing existing high nature-value habitats by alleviating their fragmentation (%)	Unknown	Quasi-experimental	National database (Natura 2000)
IT, Veneto	EXA 2007-2013	214	CMEF, change in trend in biodiversity decline, measured by bird population in agricultural areas	Unknown	Experimental	National database (MITO2000, Italian Ornithological Monitoring)
IT, Veneto	MTE 2007-2013	214	CMEF, restoration of biodiversity (FBI index)	Unknown	Experimental	National database (MITO2000, Italian Ornithological Monitoring)
IT, Puglia	MTE 2007-2013	214	CMEF, restoration of biodiversity (FBI index)	WW	Experimental	National database (MITO2000, Italian Ornithological

						Monitoring)
LT	EXP 2004-2006	214 and 221[10]	Protected or improved biodiversity	WW	Naïve	IACS
FR	MTE 2007-2013	214	CMEF, change in trend in biodiversity decline, measured by the birds population in agricultural areas	WW	Quasi-experimental	National database (PDRH, IFEN)
FR	MTE 2007-2013	214	CMEF, stemming the decline of biodiversity (FBI index)	WW	Quasi-experimental	National database (PDRH, IFEN)
MTE = Mid-term evaluation report | EXP = Ex-post evaluation report | EXA = Ex-ante evaluation report WW = With-and-without comparison						

Table 7 Counterfactual analysis in High Nature Value related impact assessment

Country	**Report**	**Measure or Programme**	**Impact indicator**	**Type of CF analysis**	**Control group formation**	**Source of data for the CF**
UK, England	MTE 2007-2013	214	CMEF, change in trend in biodiversity decline by measuring farmland species population	Unknown	Unknown	Prior literature (Natural England report, scientific literature)
IT, Veneto	MTE 2007-2013	214	CMEF, change in trend in biodiversity decline	Unknown	Experimental	National database (MITO2000, Italian Ornithological Monitoring)
IT, Veneto	MTE 2007-2013	214	CMEF, UAA area classified as HNV	WW	Quasi-experimental	National database (MITO2000, Italian Ornithological Monitoring)
FR	MTE 2007-2013	211	CMEF, UAA area classified as HNV	Unknown	Quasi-experimental	National database (PDRH, Corine Land Cover Data, IRENA)
LT	EXP 2004-2006	212 and 214[11]	Maintenance of HNV sites	WW	Naïve	IACS
FR	MTE 2007-2013	226	CMEF, maintenance of HNV agricultural and forest land	Unknown	Unknown	National database (PDRH, Corine Land Cover Data, IRENA)

[10] Same counterfactual analysis was used for both measures.
[11] Same counterfactual analysis was used for both measures.

MTE = Mid-term evaluation report | EXP = Ex-post evaluation report | EXA = Ex-ante evaluation report *WW = With-and-without comparison*

Table 8 Counterfactual analysis in soil related impact assessment

Country	Report	Measure or Programme	Impact indicator	Type of CF analysis	Control group formation	Source of data for the CF
DE, Thüringen	MTE 2007-2013	214	Annual soil loss (t/ha)	WW	Unknown	IACS
DE, Baden-Württemberg	EXP 2000-2006	121	C factor for soil erosion	WW	Quasi-experimental	IACS
DE, Lower Saxony	MTE 2007-2013	214 A2	C factor for soil erosion	WW	Quasi-experimental	IACS
DE, Lower Saxony	MTE 2007-2013	214 A7	C factor for soil erosion	WW	Quasi-experimental	IACS
DE, Lower Saxony	MTE 2007-2013	214 C	C factor for soil erosion	WW	Quasi-experimental	IACS
IT, Veneto	MTE 2007-2013	214 A, E	CMEF, maintenance / increase the organic matter content in soils	WW-BA	Quasi-experimental	IACS
IT, Veneto	MTE 2007-2013	214 B, I	CMEF, maintenance / increase the organic matter content in soils	WW-BA	Quasi-experimental	IACS
IT, Veneto	MTE 2007-2013	214 C, I	CMEF, maintenance / increase the organic matter content in soils	WW-BA	Quasi-experimental	IACS
LT	EXP 2004-2006	214	Reduced erosion	WW	Naïve	IACS
FR	MTE 2000-2006	F	Proportion of arable land under organic farming, integrated production, and pasture with less than 2 LU / ha	WW	Unknown	National database (PDRN, RICA monitoring)
FR	MTE 2007-2013	214	CMEF, maintenance / increase the organic matter content in soils	WW-BA	Unknown	National database (PDRH, Corine Land Cover Data)
MTE = Mid-term evaluation report | EXP = Ex-post evaluation report | EXA = Ex-ante evaluation report *WW = With-and-without comparison | WW-BA = With-and-without and before-and-after comparison*						

Table 9 Counterfactual analysis in landscape related impact assessment

Country	Report	Measure or Programme	Impact indicator	Type of CF analysis	Control group formation	Source of data for the CF
DE, Rheinland-Pfalz	MTE 2007-2013	214, grassland measures	Characteristic landscape	WW	Unknown	National database (FRIDA environmental database)
IT, Veneto	EXP 2000-2006	214 F 2, 3	Farmland contributing to perceptive differentiation in the landscape (number of	WW	Naïve	National database (LAND USE)

			sites and area)			
IT, Veneto	EXP 2000-2006	214 H	Additional attractive/valuable area or sites due to assistance	WW	Naïve	National database (ISTAT)
IT, Veneto	MTE 2007-2013	214	Willingness to pay for attributes of conservation and landscape	WW	Naïve	National database (case study, ISTAT)
IT, Veneto	MTE 2007-2013	323 A	Conservation and upgrading of the rural heritage	WW	Naïve	National database (ISTAT)
LT	EXP 2004-2006	212, 214 and 225[12]	Preservation of traditional landscape features	WW	Naïve	IACS
MTE = Mid-term evaluation report | EXP = Ex-post evaluation report | EXA = Ex-ante evaluation report WW = With-and-without comparison						

Table 10 Counterfactual analysis in animal welfare related impact assessment

Country	Report	Measure or Programme	Impact indicator	Type of CF analysis	Control group formation	Source of data for the CF
AT	EXP 2000-2006	121	A large set of ethological indicators (social behaviour, movement, rest and sleep, food intake, excretion, reproduction, comfort and exploration) and animal species (cattle, pigs)	BA	Quasi-experimental	National database (survey data), IACS
IT, Veneto	EXP 2000-2006	M	CMEF, I.4-2.1. Share of assisted products sold with EU-level labelling schemes	BA	Naïve	National database (ISTAT, RICA monitoring)
IT, Veneto	EXP 2000-2006	M	CMEF, I.4-2.1. Share of assisted products sold with quality label with national level labelling schemes	BA	Naïve	National database (ISTAT, RICA monitoring)
IT, Veneto	EXP 2000-2006	M	CMEF, I.4-2.1. Share of assisted products sold with other labelling schemes	BA	Naïve	National database (ISTAT, RICA monitoring)
IT, Veneto	EXP 2000-2006	M	CMEF, I.7-2.1. Share of animals on assisted holdings supported with animal welfare as a direct aim	BA	Naïve	National database (ISTAT, RICA monitoring)
IT, Veneto	EXP 2000-2006	M	CMEF, I.7-2.1. Share of animals on assisted holdings supported with positive collateral animal	BA	Naïve	National database (ISTAT, RICA

[12] Same counterfactual analysis was used for both measures.

			welfare effects			monitoring)
IT, Veneto	EXP 2000-2006	M	CMEF, I.7-2.1.Share of animals on assisted holdings supported with positive effects related to welfare standards	BA	Naïve	National database (ISTAT, RICA monitoring)
IT, Veneto	EXP 2000-2006	M	CMEF, I.7-2.1. . Share of animals on assisted holdings supported with positive effects related to EU-welfare standards	BA	Naïve	National database (ISTAT, RICA monitoring)
IT, Veneto	EXP 2000-2006	214 F	CMEF, VI.2.A-2.1. Area with beneficial lay out of crops, types of crop (including associated livestock), crop-combinations and size of uniform fields maintained/reintroduced thanks to assisted actions	WW	Naïve	National database (ISTAT, AVEPA)
IT, Veneto	MTE 2007-2013	215 1	CMEF, Animal welfare	BA	Unknown	National database (RICA monitoring), FADN
IT, Veneto	MTE 2007-2013	214 F 2.1	CMEF, Animals in danger of extinction	BA	Unknown	National database (RICA monitoring), FADN
FR	MTE 2000-2006	G	CMEF, Proportion of arable land under organic farming, integrated production, and pasture with less than 2 LU / ha	WW	Unknown	National database (PDRN)
FR	EXA 2000-2006	G	CMEF, I.4-2.1. Share of assisted products sold with EU-level labelling schemes	WW	Naïve	National database (PDRN)
FR	EXP 2000-2006	M	CMEF, I.4-2.1. Share of assisted products sold with EU-level labelling schemes	WW	Naïve	National database (PDRN)
FR	EXA 2007-2013	133	CMEF, Individual aid for quality	WW	Naïve	National database (PDRH)
FR	MTE 2007-2013	232	CMEF, Animal welfare	BA	Unknown	National database (PDRH, RICA monitoring), FADN

MTE = Mid-term evaluation report | EXP = Ex-post evaluation report | EXA = Ex-ante evaluation report
WW = With-and-without comparison | BA = Before-and-after comparison