ESPON

EUROPEAN UNION

Co-financed by the European Regional Development Fund

Inspire Policy Making with Territorial Evidence

TRANSNATIONAL BRIEF

Net impact of Interreg

An inquiry into causal inference, counterfactual and qualitative methods and their applicability for estimating the net impact of Interreg programmes

Interreg

EUROPEAN UNION

30 years together

Introduction

Net impact assessments play a role in all types of evaluation, including impact evaluation and implementation evaluation. Implementation evaluations supporting the smooth delivery of a programme are more likely to be useful in the early stages of implementation. Evaluation capturing the effects of priorities and looking into their theory of change will occur at a later stage. The legal provisions do not prevent Member States from carrying out mid-term evaluations.

The Commission encourages Member States to include, on a voluntary basis, the evaluation of the impacts of similar interventions in a previous programming period. This can make sense as for many interventions it takes years before the effects are fully realised (e.g., for large scale infrastructures, RTD projects). At least once during the programming period, an evaluation shall assess how support from the CSF Funds has contributed to the objectives of each priority. Impact evaluations according to the Commissions guidelines should be using theory based approaches or counterfactual methods are appropriate tools.

Counterfactual impact evaluation is a set of techniques borrowed from statistical and medical science and have the potential to provide a credible answer to the question "Does it work?". The central question of counterfactual evaluations is rather narrow — how much difference does a treatment make? The method then produces answers that are typically numbers, typically differences, to which it is plausible to give a causal interpretation based on empirical evidence and some assumptions. Is the difference observed in the outcome after the implementation of the intervention caused by the intervention itself, or by something else? Evaluations of this type are based on models of cause and effect and require a credible and rigorously defined counterfactual element to control for factors other than the intervention that might account for the observed change.[1]

In the Interreg context the application of such methods faces numerous challenges related e.g. to the financial size of programmes in relation to other factors. Border areas in total are home to 37,5% of the EU population. CBC programmes however receive comparably little funding (~5% of the ERDF and ~2% of all ESI-funds). CBC programmes usually have a low number of beneficiaries, thus a low number of interventions, while numerous other funds are targeting the same geographical area, oftentimes in overlapping thematic fields.

[1] EU Commission (2015): Guidance Document on Monitoring and Evaluation: Concepts and Recommendations; European Cohesion Fund, European regional development Fund; DG Regio; Brussels

What is to be done – how is it linked to Interreg

"The starting point in designing any public intervention is to identify a problem to be addressed" (EU Commission 2015a[2]). The "need-driven" approach has been the overarching principle of EU co-financed funds – with "need" being defined as observable significant difference between the status quo and a situation as it should be (need is then the gap in results – see Kaufman et al. (1999[3])) - for years.

In the reality of EU policies this identification of needs is a crucial element of policy formulation and therefore a compulsory element of any ex-ante assessment of policies[4]. However, the attachment of all policies to the overarching societal needs as defined e.g. in the form of the EU2020 strategy in practice has led to several problems.

- The necessity of linking every EU Support to the EU2020 goals led to the territorial needs analysis producing too trivial causal links, fitting the hierarchy of goals rather than the actual pressing needs
- Capturing the societal and territorial needs with appropriate indicators on the right geographical scale in many cases is not possible due to a lack territorially based information
- Due to this lack of information and the narrow scope of the needs, the similarity of policy targets and policy support strategies has led to a duplication of approaches all over Europe, which in some policy fields (e.g. innovation, RTD) has had the consequence of competition and overlaps between regions

This raises the question, at which territorial level needs can accurately be defined and subsequently impacts can be assessed by any programme evaluation. This level naturally will vary with the policy or programme in question. While for example an industrial policy will show its "footprint" at the national level or even above that, there still will be regionally dispersed effects. Environmental policies connected e.g. to water basins might show impacts not corresponding to any of the "standard" statistical regions, but rather connected to the basin area itself, cutting across multiple regions or even Member States. This means that information, which would support policy design and agenda setting, should be adjusted to the question of how much leverage a policy may achieve in the face of multiple needs:

This may be exemplified by the example of EU Regional Development Funds and their ways of addressing societal/ territorial needs. At the outset of the ongoing programming period the Commission has clearly pointed out that…

"There shall be thematic orientation – i.e. the focus on results, which wanted to be achieved (with regard on limited financial resources). The number of specific objectives per investment priority should be limited as far as possible, to facilitate the monitoring of progress and avoid the fragmentation of programmes. As a general rule programmes should set out priority axes which are mono-fund, cover only one thematic objective and a single category of region. (EU Commission 2012, 2013)

Reality showed however:

- Average number of priority axes: 7 [min.: 1, max.: 16]
- Average number of investment priorities used: 13 [min.: 0, max.: 40]
- Average number of corresponding specific objectives: 19 [min.: 1, max.: 72]
- Average number of result indicators: 24 [min.: 2, max.: 123]
- Average number of output indicators: 38 [min.: 5, max.: 163][5]

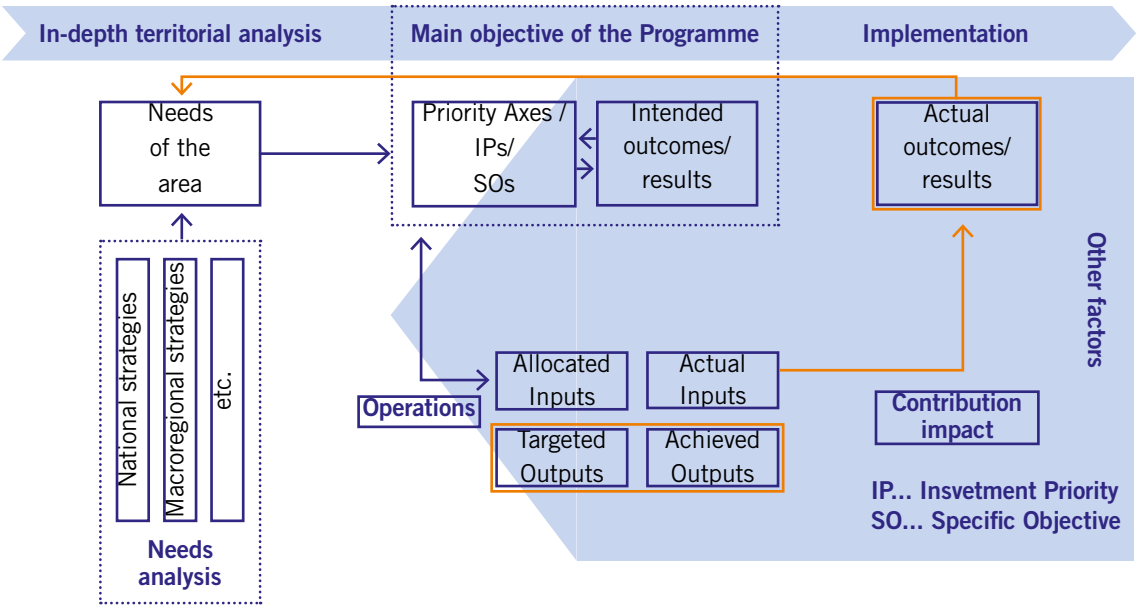
[2] European Commission (2015a): Guidance Document on Monitoring and Evaluation; European Cohesion Fund, European Regional Development Fund: Concepts and Recommendations; Brussels

[3] Kaufmann, D.; Kraay, A.; Zoido-Lobaton, P. (1999): Aggregating governance indicators ; Policy, Research working paper ; no. WPS 2195. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/167911468766840406/Aggregating-governance-indicators

[4] In the single Operational Programmes of EU policies this step is anchored in a thorough territorial analysis of the programming areas and the conducting of a SWOT

[5] results from EU Commission (2016): The implementation of the performance frameworks in 2014-2020ESI Funds; Contract: 2014CE16BAT063; Brussels; https://ec.europa.eu/regional_policy/sources/policy/how/studies_integration/impl_pf_esif_report_en.pdf

This means that the limited resources available within each programme have been largely thinned out and spread over a variety of needs addressed. However the needs described and translated into objectives have not been adjusted to the limited leverage potential of the policy as a consequence. So also in this case the major challenge is to find enough evidence for the application within the policy cycle – be it in the form of justification or publicity.



Source: OIR, 2013 based on: Barca, McCann 211: 4; European Commision, 2013a: 5
Figure 1. Programme intervention logic

What becomes clear from the figure is the fact that every policy intervention follows a clear cause-effect chain, which links the needs via the input, output through to the results, which in return address the needs and hopefully mend them in the end. – What becomes clear furthermore is the fact that policy is not happening in the void, but various other intervening factors (overall socio-economic development, other policies targeting the same territory etc.) may play a role and may influence the results of the policy and thus the achievement of the objectives (i.e. the addressing of the societal/ territorial needs). Within this world of the intervention logic information about the causal chains plays a crucial role and therefore the role of indicators is a prominent one.

The indicators applied in programmes have to establish a cause-effect chain from the need identified in the territory over the policy intervention (Input) to the direct consequence of policy (Output) through the immediate effects with reference to the direct addressees (Direct Result) to the indirect effects of the policy intervention on the objectives (as reflecting the needs) (Results). In the territorial context that means, the indicators have to establish the territorial effects of policy in the programming area (results) depicting the range of objectives (explicit & implicit) of the programme.

Results

Result indicators relate to programme objectives. They represent (in qualitative or quantitative terms) what can be measured/ captured at programme level to show contribution towards the objectives set for the programme. Result indicators are defined by baselines and targets. In other words they follow the logical chain of the questions: "What do you want to change?" - "How do you capture the change?" Thus they are strengthening the result orientation of the programming. In their design they have to be sufficiently close to policy and the societal/ territorial needs they want to address. Due to the complexity of the objectives quantitative or qualitative targets shall be combined. Regular monitoring shall prompt the policy debate and feed the self-reflection of policy.

In publications by the former head of evaluation Unit of the Directorate General for Regional development (see Gaffey, Schulte-Brauks, 2013[6]) it is recommended that result indicators shall be "selected by programmes and shall not be common indicators thus recognizing the different "journeys" to EU2020 objectives. Evaluation is to be expected to disentangle the contribution of the policy to change from the influence of other factors (impact) – thus result indicators are supposedly to be depicted as "net-effects".

Unlike outputs it becomes clear from the multitude of definitions and specifications that result indicators are the most complex issue in providing information for the policy cycle. They are supposed to build the bridge between the direct consequence of the policy and the overall effect of the interventions. It has already been pointed out in that this attribution gap may be quite substantial if the overall policy goal is formulated and captured in a very general way ("prevent demographic change", "strengthen the innovation in regions") and information deficits are then to be foreseen. But even if the attribution gap is not as large and fuzzy the quality of the result indicators should be able to capture the effects of policy in a way, that excludes all potential other intervening factors:

While the first issue of closing the attribution gap may be covered by a variety of different indicators (qualitative and quantitative), the latter problem (i.e. the assessment of the "net-effect" may be tackled by the way information is combined (see section below).

The set of common indicators as foreseen in the annexes to the ERDF, Cohesion Fund and ETC regulations[7] is designed to fit programmes all across the EU, however in practice suffers from multiple shortcomings. The main challenges for all these indicators are on the one hand the harmonisation of definitions (e.g. when is a product "new" to the firm?) and the collection of data within the programme. The suggestions from the legislative proposal for the upcoming programming period post 2020[8] foresee the introduction of so called direct result indicators, which will be strongly linked to the performance of the single operation/ measure. – They strongly resemble the character of outputs in this way and certainly will not close the attribution gap between the policy action and the overall need addressed by the programme.

[6] Gaffey, Veronica; Schulte-Brauks, Antonella (2013): Results and Performance for Cohesion Policy. A Step Change for the Future. Vilnius, 4 July 2013

[7] Regulation (EU) No 1301/2013 of the European Parliament and of the Council of 17 December 2013 on the European Regional Development Fund and on specific provisions concerning the Investment for growth and jobs goal and repealing Regulation (EC) No 1080/2006

Regulation (EU) No 1302/2013 of the European Parliament and of the Council of 17 December 2013 amending Regulation (EC) No 1082/2006 on a European grouping of territorial cooperation (EGTC) as regards the clarification, simplification and improvement of the establishment and functioning of such groupings

Regulation (EU) No 1300/2013 of the European Parliament and of the Council of 17 December 2013 on the Cohesion Fund and repealing Council Regulation (EC) No 1084/2006

[8] see e.g. EU Commission (2018): ANNEX to the Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Regional Development Fund and on the Cohesion Fund; COM(2018) 372 final; Strasbourg; https://eur-lex.europa.eu/resource.html?uri=cellar:8d2f7140-6375-11e8-ab9c-01aa75ed71a1.0001.02/DOC_2&format=PDF

The approaches for assessing net-effects of policies

The last element of the use of information in the policy cycle is the ongoing and ex-post observation of the policy effects and the judgement on the success/ failure of the intervention - i.e. the evaluation. This exercise is of high relevance for setting up future programmes as the learning loop of identifying what worked and what did not allows for a subsequent adjustment of future programmes in order to achieve better results.

Comparability and potential for aggregation

'When you can measure what you are speaking about, and express it in numbers, you know something about it; when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.' Lord Kelvin (cited in Samuelson and Nordhaus, 1989[9]) has pointed out quite accurately the basic problem connected to this part of evaluations – i.e. the difficulty of finding operational measures for such complex concepts as utility and the full value of something – let alone the problem of expressing these findings in a quantitative way – thus making it compatible with comparisons in the next step. When assessing the value of something the major challenge is to transform as much information as possible from reality into a form which is accessible for our sensitive apparatus (i.e. computing capacity, senses).

Comparison

This is the central task of evaluation. It is often forgotten and/or done implicitly, thus leading to major biases when interpreting the final results. Still, when talking about comparisons of facts and measures, four major philosophical concepts have to be taken into account (see e.g. Martinez-Alier et al., 1997[10]):

- Strong commensurability, according to which there exists a common measure of the different consequences of an action, based on a cardinal scale of measurement.
- Weak commensurability, according to which there exists a common measure based on an ordinal scale of measurement.
- Strong comparability, according to which there exists a single comparative term by which all different consequences can be ranked.
- Weak comparability, according to which values are irreducibly plural and cannot be uniquely ordered along a single scale.

Creating comparability and thus aggregating results to an overall evaluation can be achieved by the use of suitable methods of aggregation in the light of the multitude of indicators which will be available side-by-side to express success/ failure of policy. First it has to be assumed that the single indicators are already depicting net effects (see next section), then aggregation methods should be applied, which would allow for a maximum of transparency when comparability is produced.

The possibility of isolating the "pure", causal effect of the policy

In many cases this is not an easy task, because several external and confounding factors may influence the selected result indicators. How is it possible to isolate the change in the result indicator generated by the policy? In an ideal situation, we would like to compare the change in the result indicator in a world where the policy was implemented against a world where the policy was not implemented (and identical for all the rest). The difference between the two values would be an accurate measure of the direct effect of the policy. Obviously, this is not possible. Nevertheless, the difference-in-difference (DID) method allows overcoming this issue[11].

The principle behind the DID method can be explained by an example from the medical field. Consider the case of testing a new medical treatment. The reaction to this treatment depends on many

[9] Samuelson, P.A. and W.D. Nordhaus (1989), Economics, 13th edn, New York: McGraw-Hill.

[10] Martinez-Alier, J., G. Munda and J. O'Neill (1997), *'Incommensurability of Values in Ecological Economics'*, in M. O'Connor and C. Spash (eds), *Valuation and the Environment – Theory, Method and Practice*, Cheltenham, UK and Lyme, USA: Edward Elgar.

characteristics of the patients, like age, weight, etc. For this reason, tests of this kind are often performed on couples of twins: one receiving the treatment and the other not. Since couples of twins are rare, the same test can be performed on two individuals as similar as possible in the relevant characteristics (age, weigh, etc.). If we find the perfect match of our treated patient, we can compare the two individuals and measure the effect of the treatment.

The application of DID to the policy evaluation follows the same logic. Imagine an example about the policy aimed at promoting product and process innovation in the field of CO2 reduction and sustainable energy. The treatment, in this case, is represented by the policy itself. The treated units (for instance the firms receiving funding) must be matched with the most similar untreated units (i.e. that did not receive the policy). "Similarity" is defined based on all the external factors potentially having an effect on the policy results. For example, suppose that firms located in urbanized regions are more likely to undertake innovation activities, because the customers living in these areas are more sensitive to environmental issues than those living in rural settings. If it is the case, we should match treated and untreated units based (among other things) also on their location (urban vs rural). Assume, for instance, that the implementation of innovation is also depending on the sector in which firms operate. Therefore, the matching should associate units pertaining to the same sector.

The next step after the matching procedure is represented by the comparison of treated and untreated units similar between them. Recalling the medical example above, the matching is a procedure for identifying sort of "twin-units" (treated and untreated) that can be therefore compared in order to identify the pure effect of the treatment, i.e. of the policy.

Finding comparable units in the context of any programme nonetheless might prove a difficult task. Several approaches relying on statistical methods have been developed.

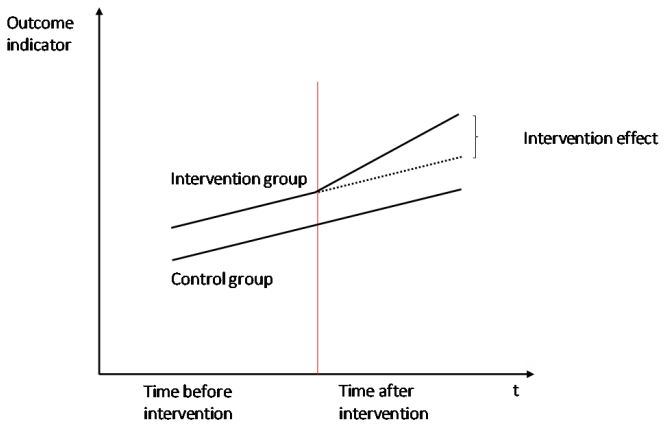


Figure 2. Difference in differences

Applicability within Interreg

As statistical methods for net-impact calculations in principle rely on a certain (depending on the method and the circumstances) minimum number of observed cases, Interreg programmes with their inherently limited numbers of projects already suffer some difficulties from the outset. In many programmes under the current circumstances, the overall number of projects is already too low to allow for statistically significant results. However even when there are sufficient case numbers in some larger programmes, the thematic differentiation might not allow for identifying enough cases which are actually comparable. Some of the approaches as described below can thus be considered rather as methodological best practice, which for application in Interreg need careful preparation from the programme setup onward.

[12] Observable traits of the groups are determined by the selection criteria of the policy measure to be evaluated (e.g. specific income or lack thereof, specific education attainment level, company size in terms of employees or turnover etc.)

Unobservable traits are elements of the character of the (potential) beneficiaries/ objects of evaluation, which are not specifically targeted but which nonetheless influence the funding/ non-funding decision, propensity to apply for support (e.g. cultural background, upbringing and values, religious background, etc.)

Randomized Sampling (the statistical gold standard)

This approach corresponds to the random selection of beneficiaries (intervention group) and non-beneficiaries (control group) from the eligible population in two groups in advance of the policy implementation. The rationale is that only by being able to select both – test and control groups – from a randomized and large enough sample size any differences in observable and unobservable traits[12] of the two groups may be minimized. Especially the unobservable characteristics of the evaluation objects may only be eliminated with this approach.

Over the course of the policy implementation, the development of these groups can be observed and their resulting differences considered as the net-impact of the programme.

Applicability within Interreg

A reliable application within Interreg is uncertain, notably due to limited numbers of beneficiaries and broad thematic orientation. Large Randomised Control Trials (RCTs) are time-consuming and costly, although an acceptable level of accuracy can be attained with existing data. Furthermore, selection criteria for Interreg subsidies consider observable traits of applicants and may cause a non-randomized approach if these observable traits are influenced by unobserved confounders. This necessitates prior analysis as to whether selection criteria are defined well enough so as to regard the unobservable traits as only minimally influencing the support/ non-support situation.

Propensity Score Matching (PSM)

Statistical matching seeks to minimize the selection bias and attain an accurate estimate of the Average Treatment Effect (ATE). This is done by deducting the mean of a certain indicator of non-treated units (i.e. non-beneficiaries) from the mean of the treated units (i.e. beneficiaries). The result, however, contains the ATE and a selection bias, i.e. the treated units have not been randomly selected but certain observable traits suggest that they actually necessitate the treatment (also known as reversed causality).

Matching is assumed to solve the problem of selection bias and reversed causality. The idea is to match units within a test and control group and that the only difference between the matches is the fact of the treatment (i.e. beneficiaries vs. non-beneficiaries). It is a statistical approach that identifies in a group of non-participants those units who are similar to the participants in the treatment in all relevant pre-treatment characteristics. The simplest method of matching compares units (i.e. beneficiaries) along important observable characteristics and then matches very similar units to pairs. It is assumed that the measurable outcome is independent of the selection for treatment, conditional on these observable characteristics, also known as the Conditional Independence Assumption. This reduces the selection bias and a reliably accurate ATE can be obtained.

Propensity score is the probability of participation in the treatment (i.e. EU-funded project) estimated based on multiple observed traits that treated and non-treated units have in common. The different scores are obtained by a logistic regression and divided in strata. Treated units can then be accurately matched with counterfactual units (i.e. non-treated) based on the respective strata. ATE can be obtained for every strata, and the overall ATE is then obtained by the weighted average of the ATE obtained within individual strata. In other words, this is the difference between the outcomes after participation in a programme and non-participation.

The crucial difference of PSM from conventional matching is that it matches subjects on one score rather than multiple variables. A necessary assumption for this method though is, that all relevant differences between beneficiaries and non-beneficiaries are captured by observable variables. An issue for matching in general is the richness of the dataset and/or availability of appropriate matching variables which should be chosen on the basis of economic (and also other relevant theories).

Applicability within Interreg

The law of large numbers applies, where obtaining efficient and consistent estimates of ATE necessitates a large enough totality of units. Most examples of matching approaches in policy evaluation are therefore to be found in large-scale policy support measures with a high number of homogenous beneficiaries (e.g. SMEs, education and training measures for low qualified workforce etc.). In Interreg, there is hardly any measure, which supports a high number of homogenous beneficiaries.

Regression Discontinuity Design (RDD)

This method can be applied if the eligibility for a programme is determined by a rule of the following type: those above a certain threshold are eligible for the programme while those below are not eligible, or vice versa. Threshold examples include age for people, income for households and number of employees for firms. In the neighbourhood of the threshold we have a situation that resembles randomization: the units around the threshold receive sharply different treatments, despite having similar values for the selection variable (Fig. 3).

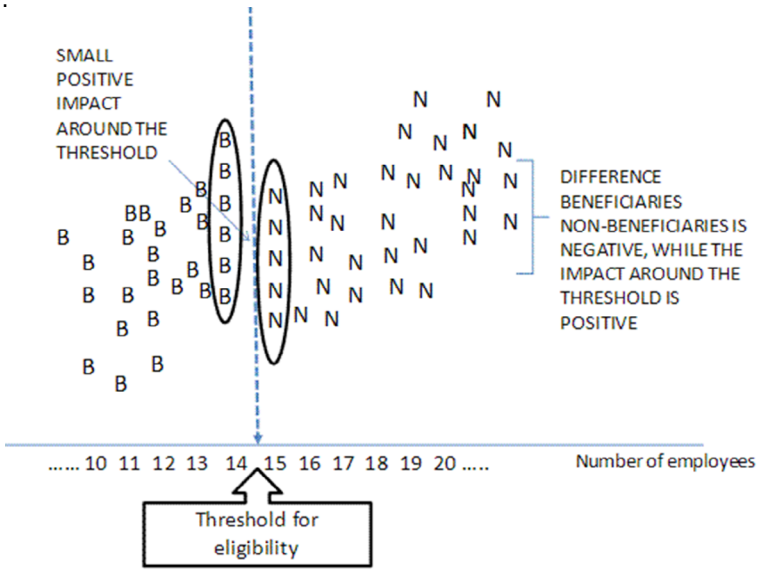


Figure 3. Regression disocontinuity

Firms to the right of the threshold (the N's) do not have access to the funding, and this rule is strictly enforced. The context being, that larger firms tend to perform better on the performance measure, and smaller firms are targeted for the funding. Thus, the intervention has a clear compensatory rationale; it targets the less advantaged fraction of the population. The results of naive comparisons in this case could be misleading: if one were to take all the N's and all the B's and compare their average performances, one would conclude that the funding lowers firms' performance. We know that such causal interpretation is unwarranted because the two groups are different even in the absence of the intervention (and there is negative selection bias).

The RDD method does not suffer from selection bias: if we restrict the attention to marginal individuals and compare firms with 14 employees (marginally eligible) to firms with15 employees (marginally ineligible), it seems reasonable to assume that a difference of 1 employee will have a minor impact on the performance measure. By contrast, the two subgroups are treated very differently by the programme, so that around the threshold we have a randomization of sorts.

The difference between the average performances of the two samples of 14-employee firms and 15-employee firms is a credible estimate of the impact of the programme. In the example, the impact around the threshold is positive, indicating that the programme works. Visually, the impact is represented by the "downward slide" of the N's just above the threshold with respect to the B's right below the threshold.

Applicability within Interreg

With this approach, the law of large numbers will not be as stringently needed in order to identify the two groups for comparison (funded and non-funded). In the Interreg context this would call for a measure, which clearly cuts off the eligibility at a threshold, which does not actively determine the performance of the beneficiaries towards the objective of the intervention (e.g. the size of an institution).

So, one application case may be the support of small and medium sized towns across borders with respect to developing land use plans. If the eligibility criterion is e.g. a certain town size (number of inhabitants), the establishment of test and control groups may be conducted straight at this criterion selecting non-funded towns (who also develop land use plans) which are just above the eligibility threshold and compare their performance (e.g. with respect to sustainable land use) with the funded ones thus establishing a DiD causal inference.

Pipeline approach

The pipeline approach takes as the comparison group individuals, firms or communities which have been selected to participate in the programme but have not yet done so. When a project is implemented in phases, subjects in the second phase (i.e. who will not receive benefits until some later point in time) can be used as the control group for Phase 1 subjects. Communities, people or firms selected for project participation, but not yet treated, are chosen for the control. Since they have also been selected for treatment there should in principle be no selection bias. For this assumption to hold, one shall expect no changes in selection criteria and no ranking of beneficiaries of both phases. However, policy design and selection criteria must be carefully reviewed when applying this approach because there will often be systematic differences between the phases.

Applicability within Interreg

Typical applications are for example training programmes, enterprise support, etc. where the finances dried up early in the programming period. Application on Interreg is possible as this matching approach does not necessarily rely on large numbers for causal inference. It may be applied where the conditions of at least two calls do not change over time, i.e. eligibility criteria, budget volume, selection process). The two consecutive cohorts of two calls may then be used for the distinction of the treated and non-treated groups. The first group, who has already shown performance after the support of the Interreg programme will then be compared with the performance of the group which has already been selected by the next call but has not yet been funded. Yet again the performance of the groups will have to be independent from the selection criteria.

Pragmatic matching on case by case basis

While statistical approaches are considered the "gold standard" in terms of quantitative net-impact determination, the reality of Interreg programmes oftentimes calls for more "pragmatic" approaches, like those applied in the context of the ESPON projects like Ex-post territorial impact assessment of cross-border cooperation programmes (ESPON TIA CBC) and Territorial evidence support for European Territorial Cooperation programmes (ESPON TEVI).

The “small scale counterfactual” is a qualitative method analogous to DID but is less time and resource consuming than DID. It involves calculating the net-impact of the programme by comparing the actual development of a region`s values for a given indicator with a hypothetical scenario in which no actions have been taken by the programme in the region.

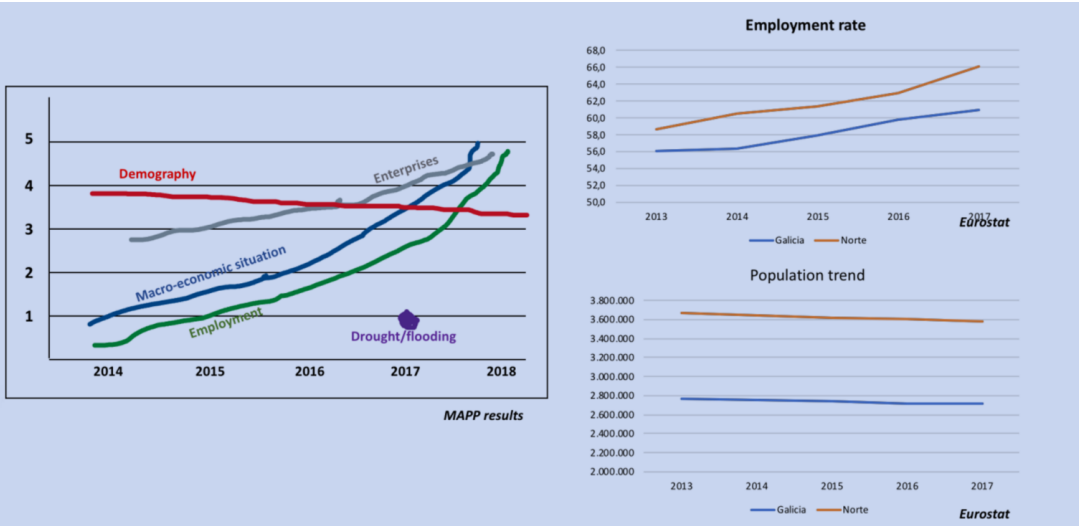
Unlike the proper counterfactual approach the test- and control groups are in this case not established through statistical matching methods (e.g. propensity scores, discontinuity- or pipeline approaches) but on a case-by-case selection matching funded with non-funding entities which show the same observable traits (i.e. qualities as expressed by the selection criteria of the measures which are to be assessed). This “small-scale” approach is justified by the fact that both test and control groups will be too small in reality to establish statistically sound matching methods; thus, it seems justified to compare in a DID assessment the changes over time of both the treated with the non-treated cases, which will provide a net effect of the assessed measure within the programme. As a result, this simplified method involves establishing a group of beneficiaries and a group of non-beneficiaries who are active in the same fields, enabling a comparison between the two groups. Data on the indicator(s) in question and the contribution of the groups to that has to be obtainable.

With two groups available, the change in value of respective indicators is compared and the difference between the two values represents the net effect of the programme.

Funding framework approach

If indicators in question typically rely on various funding resources, the establishment of the funding framework for a region, via MAPP, can be used to identify the net effect. MAPP is an approach initially developed by Dr. Susanne Neubert at the German Development Institute . It allows assessing impacts of interventions basing on three main elements: life curve, trend analysis, and influence matrix. The life curve sets the context for the assessment, the trend analysis shows the overall trends of different indicators (i.e. irrespective of any specific programme), while the influence matrix, basing on the two previous elements, constitutes the assessment of net effects based on the funding framework involved in the region. All of these tools use a point system (from 1 to 4) and are based on expert judgement. MAPP, thus, should be ideally conducted in a workshop setting.

The Life curve shows the overall development trends (based on indicators to be selected by the group, e.g. employment) in the cooperation area along a certain time-frame, beginning before the programme started and ending at present. Participants should be asked to assess the development of each indicator each year according to a five point scale. These assessments can be supported by data on such indicators if available.



Source: ESPON CBC TIA, 2019.

Figure 4. Life curve produced in ESPON TIA CBC project

In the *Trend analysis matrix,* detailed development trends of the selected indicators are assessed over the same time period. Workshop participants should be asked to score each indicator from 1 to 4 for every year and for every region, giving a general trend from the first to the last year as a gross magnitude. The regions for that purpose have to be defined by the participants, i.e. in cases of programmes with NUTS3 division, any deviating administrative units (e.g. functional areas) are defined by the participants (Fig. 5).

Trend analysis	Year Trend					
	2014	2015	2016	2017	2018	201-18
Improve the participation of the business sector in innovation processes and R+D+i activities closer to the market (1B)						
No of companies that cooperate with research centres	1	1	2	3	3	+
Joint projects developed between enterprises and institutions	1	2	2	3	3	++
No of beneficiary companies that introduce new products for the company	1	1	1	2	2	+
Increased number of enterprises that have invested in R+D+i	1	1	1	2	2	+
Size of investments by companies in R+D+i	1	1	2	3	2	+
Improve the necessary and favourable conditions for the appearance of new business initiatives (3A)						
No of services for enterprise development created or supported	1	1	2	2	2	+
SME/companies with cross-border business	2	2	3	3	3	+
Enterprises created/improved in the cooperation space (of which by young/unemployed/social economy)	2	2	3	3	3	+
Companies that offer professional internships	1	1	2	2	2	+

Source: ESPON CBC TIA, 2019.

Figure 5. Example of a trend analysis as applied by the ESPON TIA CBC project

The *influence matrix* represents the net impact determination, assessing the programme against other factors influencing the development of an indicator, with consideration of the background provided by the live curve and trend analysis. These can be other funding programmes (EU, national or private) as well as non-funding related developments. This method can be used either for qualitative assessments (where the influence value is taken into account when making the magnitude judgement from the trend analysis) or for semi-qualitative assessments (where the influence value is multiplied with the gross development). The development of the influence matrix has to be done in close cooperation with actors who have best understanding of other funding schemes and relevant influencing factors in the programme area.

Influence matrix	CBC	ERDF	EAFRD	National	Others
Size of investments by companies in R+D+I	5%	40%	0%	25%	20%
Joint products related to historic, cultural and natural heritage developed	10%	35%	35%	0%	20%

Source: ESPON CBC TIA, 2019.

Figure 6. Example of an influence matrix as applied by the ESPON TIA CBC project

For given indicator, the total amount of funding available to the region is determined in the influence matrix. The share of the programme then constitutes the share of the gross impact of the programme. If no other funding scheme is available and the impact is likely to be based on funding for the most part, the gross impact can be considered as the net impact.

The problem with this approach is likely to be the data availability on a regional level. While for some funding schemes this is readily available, data availability and quality might vary significantly between countries. Additionally, the scope of the funding would have to be matched with the programme, which will only be possible for certain kinds of indicators.

Main messages

Assessing the net impact of any programme with statistical significance is a complex and oftentimes costly endeavour. Especially in the case of Interreg programmes with the numerous limitations ranging from relatively small numbers of projects through to thematic spread and limited funds available, such approaches are oftentimes neglected in favour of purely qualitative methods. Nonetheless, there are several quantitative and hybrid approaches that can be applied in Interreg impact assessments provided that the following preparatory steps are considered already in the programme intervention logic.

- Territorial heterogeneity of the programme area can lead to different impacts in different territorial contexts. While the Average Treatment Effect is assessed at programme level, regional fixed effects can be obtained as well. Thus, result indicators can reflect expected change through the cooperation programme at regional level and be aggregated as to target values at programme level.
- Fragmentation and heterogeneity of selected Priority Axes and Specific Objectives add to the difficulties of obtaining statistically significant results.
- Result indicators have to be countable (discrete numbers) or measurable (continuous numbers) rather than nominal (e.g. status of or performance of).
- Data collection has to commence with the programme setup. If e.g. characteristics of applicants, beneficiaries or desirable end-users are considered relevant for the foreseen impact assessment, these characteristics have to be recorded from the beginning of the programme.
- Impact assessment activities require respective capacities from programme authorities, notably for primary data collection.
- Cooperation with institutions possessing relevant data such as statistical offices has to be established early in the programme

ESPON

EUROPEAN UNION

Co-financed by the European Regional Development Fund

Inspire Policy Making with Territorial Evidence

ESPON 2020

ESPON EGTC
4 rue Erasme, L-1468 Luxembourg
Grand Duchy of Luxembourg
Phone: +352 20 600 280
Email: info@espon.eu
www.espon.eu

The ESPON EGTC is the Single Beneficiary of the ESPON 2020 Cooperation Programme. The Single Operation within the programme is implemented by the ESPON EGTC and co-financed by the European Regional Development Fund, the EU Member States and the Partner States, Iceland, Liechtenstein, Norway and Switzerland.

ISBN: 978-2-919795-18-5

Editor: Vassilen Iotzov, ESPON EGTC;
Authors: B. Schuh, R. Gaugitsch – ÖIR GmbH – Austrian Institute for Regional Studies
Editorial team: INOVA+

Published in June 2020