European
Commission

Counterfactual impact evaluation of European Social Fund interventions in practice

Guidance document for Managing Authorities

Social Europe

Counterfactual impact evaluation of European Social Fund interventions in practice

Guidance document for Managing Authorities

European Commission
Directorate-General for Employment Social Affairs and Inclusion
Directorate G

This guidance material has been produced based on the work by the following experts: Claudio Deiana and Giulia Santangelo, Centre for Research on Impact Evaluation (CRIE), Joint Research Centre (JRC) and the European Commission

The authors are grateful to Linda Adamaite, Sara Flisi and Paolo Paruolo for valuable comments and suggestions.

LEGAL NOTICE

Manuscript completed in May 2020

Luxembourg: Publications Office of the European Union, 2020

PDF ISBN 978-92-76-19922-9 doi:10.2767/55495 KE-02-20-333-EN-N

Contents

Introduction

The European Social Fund (ESF) is Europe's main instrument for supporting jobs, helping people get better jobs and ensuring fairer job opportunities for all EU citizens. The EU distributes ESF funding to the Member States and regions to finance employment, social inclusion and education-related projects. Designated national and regional **managing authorities** select the projects, monitor their implementation and evaluate the extent to which the aims of the ESF are met.

The aim of this document is to present **four case studies** on ESF evaluation to showcase real problems and solutions when dealing with ESF evaluations. The examples are fictitious, inspired by real evaluations and adapted for the purpose of this guidance in order to clearly illustrate typical challenges that an evaluator or managing authority can come across. The practice solutions presented here are sometimes not the optimal ones, as discussed in the text.

The four cases highlight different methodologies of a **counterfactual impact evaluation** (CIE) whose aim is to assess the effects of policy/programmes/initiatives[2] with respect to a reference counterfactual situation and to verify whether these interventions change the behaviour/conditions of a target population in the desired direction for a number of outcomes of interest. The typical evaluation questions asked are of the type: *what would have happened in the absence of the policy or intervention?* Results may help in designing policy in future to improve its effectiveness and efficiency.

Each case is presented according to the following scheme:

1. policy context
2. policy objectives
3. outcomes of interest
4. evaluation plan
5. data description
6. methodology
7. interpretation of results
8. lessons learnt.

List of examples

1) Upskilling programme
   Example of regression discontinuity design method
2) Promoting employment and mobility for the long-term unemployed
   Example of difference-in-difference method
3) Internship programme for the older unemployed
   Example of the propensity score matching and coarsened exact matching methods
4) Promotion of enterprises
   Example of the propensity score matching method.

[2] The following terms are used as synonyms for the purpose of this document:

- intervention, policy, and programme
- participant and treated unit
- non-participant, untreated unit, control unit.

1. Example of regression discontinuity design method: Upskilling programme

1.1 Policy context

The managing authority of country A designed a programme to improve workers' skills and better match demands on the labour market by providing upskilling opportunities **for young people**. The programme was implemented under the supervision of the Ministry of Education in collaboration with the Ministry of Labour, and was called the 'Upskilling programme'.

The programme specifically targeted unemployed low-skilled young individuals, i.e. those aged under 30 and with an education level lower than upper secondary. Since inhabitants of poor areas are normally less likely to have the resources to attend requalification courses, in order to define the **areas of intervention**, the managing authority of country A exploited the difference in the average income per capita at regional level. In particular, it defined those regions lying below the bottom 10th percentile of the national income distribution as areas of intervention.

The instruments included upskilling qualification courses lasting 9 months.

1.2 Policy objectives

The 'Upskilling programme' offered different instruments to acquire marketable qualifications that could **improve the labour market prospects of treated individuals**. The target population consisted of young unemployed low-skilled individuals living in low-income regions.

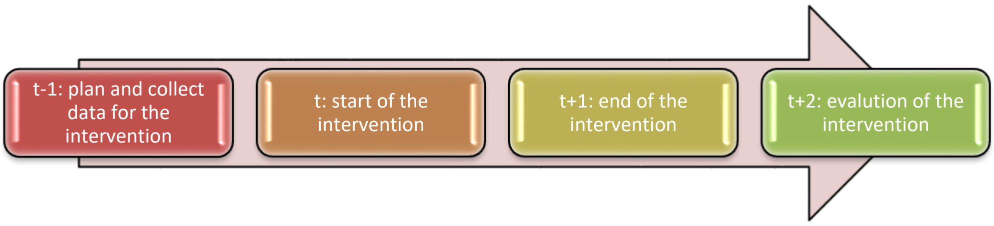
In the long run this investment was also expected to **produce a positive impact on country A's productivity and competiveness** by ensuring that the local labour market could rely on the supply of qualified workers in the near future.

1.3 Outcomes of interest

Given the policy objectives, the potential outcomes of interest include the following:

1. **labour market status** of young treated individuals after completion of the upskilling programme to determine whether they found a job;
2. productivity and competitiveness of firms in the local labour market over the medium / long term.

Figure 1: Plan for evaluation



1.4 Evaluation plan

The managing authority planned in advance what data needed to be collected for the evaluation. This phase took place in the pre-intervention period (indicated in Figure 1 as t-1). This allowed the assessment exercise's smooth implementation via CIE in the post-intervention period (indicated as t+2).

This evaluation focused on the first outcome of interest, i.e. labour market status of young treated individuals, by comparing in the post-intervention period the employment status of the following two groups:

- young unemployed individuals under 30 living in eligible regions;
- young unemployed individuals living in the same regions but not admitted to the qualification activities because they were older than 30.

1.5 Data description

The data needed for this evaluation were collected from **two administrative sources**, namely **education and employment registries**.

The **Ministry of Education** provided information about the upskilling courses attended by the young people participating in the programme. With regard to control variables, for all unemployed people in eligible regions both below (participants) and above (non-participants) the age threshold of 30, the local offices of the ministry in the regions involved provided data on demographic characteristics, level of education, grades of last year of compulsory school, and information on possible other courses attended.

As for outcome variables, data made available by the **Ministry of Labour** was used to gather information on the intervention's impact on access to the labour market and working conditions. The latter were analysed using different measures: employment status and – for those who found a job – earnings and type of contract.

The evaluator planned the collection of data from different data sources in advance. The data collection therefore started in t-1 and was done at **disaggregated level**, i.e. for individuals. The evaluation relied on merging data sources from different data holder institutions; this meant a common identifier had to be used. The data collection and merging process was carried out without obstacles because both ministries used the fiscal codes as unique individual identifiers.

1.6 Methodology

This programme represented an interesting case study where **the administrative rules** introduced since the intervention's launch provided a **clear definition of the process for selecting participants**.

This setting is suitable for applying the **regression discontinuity design (RDD)**, a 'quasi-experimental' approach relying on the fact that a measurable variable determines the selection of participants through a threshold value, which defines their eligibility for the intervention

(Cattaneo, Idrobo, and Titiunik, 2019). Around this threshold, i.e. locally, the assignment to the intervention can therefore be considered as random, making it possible to estimate the local average treatment effect (LATE).

In this particular case, the eligibility of young unemployed individuals was determined by the age threshold of 30, and people under (treated) and over (controls) 30 years old could be considered locally randomly assigned to the treatment. To make sure that the control units older than 30 provided a good counterfactual for treated units, the analysis was restricted to individuals around the threshold of 30 in the age interval 27-33.

The similarity of eligible and non-eligible young people (younger and older than 30) was also checked by comparing the characteristics of individuals in the two groups and their distributions.

1.7 Interpretation of results

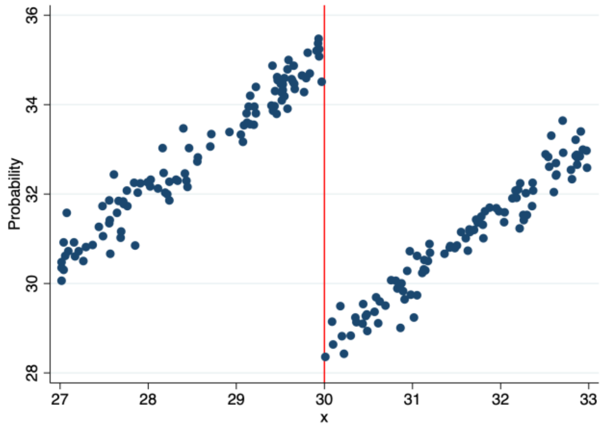
The probability of participating in a programme is by definition higher for eligible individuals than for non-eligible individuals. In the presence of a clear-cut eligibility threshold, it is discontinuous at the threshold point, represented in this case by the age 30. The size of the discontinuity in probability between the two groups of eligible and non-eligible can be measured by observing the values of the probability for units below and above the eligibility threshold and calculating the difference between the two numbers.

Figure 2 below shows the changes in the labour market status of young unemployed individuals, for the two groups of young people below and above the age threshold of 30 (indicated by the red vertical line). The programme effect is measured through the estimated LATE. This is calculated through the difference between the estimated treatment effects for the two groups of individuals below and above the eligibility threshold.

In this case, the LATE corresponds to the estimate **of the programme impact on employment status** 6 months after the end of the courses. After completion of the programme, those treated (individuals younger than 30 who participated in the programme, represented by dots on the left side of the red vertical line) had a higher probability of having a job than the controls (individuals older than 30 who did not participate in the programme, represented by dots on the right side of the red vertical line). This positive impact is detected through the jump in the probability of having a job of around 6.5 percentage points (probability of 35% for younger treated individuals vs probability of 28.5% for older untreated individuals) for individuals around the age threshold of 30 years. This estimated impact is statistically significant, i.e. it is judged not to be due to chance at a given significance level, typically taken as 5%. The significance level represents the probability of observing a large estimated effect when the true effect is 0.

Among the employed individuals, those attending the upskilling courses were on average more likely to have better paid jobs than non-participants.

Figure 2. Probability of finding a job 6 months after the end of the programme



Caption: The probability of finding a job (%) is calculated for different ages in the interval 27-33.

1.8 Lessons learnt

Estimated treatment effect

This evaluation implements the RDD approach to detect the causal effect of the 'Upskilling programme'. It should be noted that the causal effect found, i.e. the LATE, is local at the threshold point of 30 years.

Selection issue

In general, estimated treatment effects can suffer from bias due to the process of selection for the treatment. This is normally referred to as 'selection bias' in the evaluation literature and refers to the fact that individuals who decide to participate in a given intervention (self-selection) are by definition different from those who decide not to participate. If these pre-existing differences between participants and non-participants can influence the outcome of the treatment, the simple comparison of the outcomes of the two groups does not provide a correct estimate of the treatment's effect.

In this case, the possible selection bias was limited thanks to the presence of a clear administrative rule defining the selection of participants for the programme. First, the rule was set on the basis of the economic situation of regions in country A, with the selection of regions characterised by worse economic conditions compared to the national average income. Moreover, due to budget constraints, the policy targeted a specific disadvantaged subpopulation, identified in the young low-skilled unemployed. The presence of the age threshold helped to implement the CIE methods, making it possible to apply the RDD approach.

Alternative approach

As an alternative to the RDD approach, in order to estimate the treatment's effect on a wider pool of treated and control units and have a non-local estimate, the **difference-in-difference (DiD)** approach could have been applied. In this setting, one could consider the regions where the programme was not implemented to construct the counterfactual scenario. However, this method could not be applied due to data limitations. In particular, for non-eligible regions, information on similar courses (provided not by the programme but by regional institutes) was not available. Hence, it was not possible to inspect the comparability of eligible and non-eligible regions as a preliminary check for the analysis.

Data availability

Access to data from both ministries was crucial to estimate the intervention's impact. In addition, the analysis could be performed at individual level thanks to the availability of individual data from the Ministry of Education.

If available, additional data on the outcome of interest, making it possible to follow individuals in different points in time (for instance, 3, 9 and 12 months after the end of the programme) could have been used to improve the analysis, assessing possible changes in the impact over time.

2. Example of difference-in-difference method: Promoting employment and mobility for the long-term unemployed

2.1 Policy context

In the 2007-2013 programming period, country B set up a **training programme for the long-term unemployed** for the specific purpose of **supporting the transition to work** and the wider **goal of preventing episodes of poverty, crime, and social exclusion**.

The intervention was implemented by the **Ministry of Labour** and specifically targeted those regions with more than 10% long-term unemployed in the working-age population (15-64 years old).

Accordingly, the training programme was only implemented in 5 of the 10 regions in country B. Within the regions targeted, only the long-term unemployed, i.e. people who were out of work and had been actively seeking employment for at least a year, were **eligible** for the training.

The intervention consisted of **six-month training courses** covering the following topics: sewing, commercial cleaning, furniture upholstery, and catering.

2.2 Policy objectives

The short-term policy objective was to **support the transition to work** for the long-term unemployed. In addition, for the medium/long run, the policy objective included the wider **goal of preventing episodes of poverty, crime, and social exclusion**.

The initiative was designed in close relationship with local businesses and organisations in order for the long-term unemployed to acquire more skills and competencies in sectors with higher labour demand in the local markets. Importantly, no similar programmes were implemented in the non-targeted regions.

2.3 Outcomes of interest

The intervention was intended to improve the labour market prospects of the long-term unemployed by increasing their probability of being employed after attending the training programme.

Given the policy objectives, the main outcomes of interest were the:

1. **labour market status of treated long-term unemployed** after completing the training programme to determine whether they exited unemployment and entered employment;
2. **levels of poverty, crime, and social exclusion** in the regions over the longer run.

The evaluation focused on the first type of outcomes, analysing the rate of outflow from unemployment to employment in country B regions.

2.4 Evaluation plan

Implications of the selection process

Participation in the training was mandatory for all long-term unemployed claiming unemployment benefits. This eliminated the ‘self-selection issue’ normally arising for the evaluation when individuals voluntarily decide to participate in a programme. This ‘selection bias’ commonly implies that there may be **pre-existing differences** between the individuals participating and not participating in a programme.

The methodology discussed below can be used also to deal with cases where ‘selection bias’ is in place, since it allows the evaluator to control for potential differences or unobserved factors which are constant over time.

Timing of data collection

At the time of the intervention’s launch, the Ministry of Labour of country B collected **data on the treated group in the pre-intervention period**, and in particular information on demographic characteristics, job history, periods of unemployment and education attainment.

In contrast, the same **data for the control group** was **collected later** in the process since its use for evaluation purposes was not planned at the time of the policy start (when the pre-intervention information on the treated group listed above was gathered). This represented **a big challenge,** since the data extraction process had to be performed twice, doubling the work and opening the door for data inconsistencies between the two data extractions.

2.5 Data description

Data for the evaluation, for both the treated and the control group, were provided by the **employment registry**.

Data on employment status were gathered for the pre-intervention period to observe the labour market history of individuals who constitute the treatment and control groups before the programme was implemented. Pre-intervention, individual-level information was linked to information on the post-intervention period, capturing the monthly employment status of treated and control units up to 24 months after the intervention’s completion. The individual history, both pre- and post-intervention, was built on the spells of employment and unemployment registered in the database.

Since some gaps in information for some particular months emerged in the employment registry, the **social security** database was used to fill in these gaps and reconstruct the entire pre-intervention period considered.

Besides information on labour market history, the employment registry also provided demographic data on treated and control units, such as gender, level of education, province of residence, and age.

The employment history and the province of residence data were used to construct the control group and check the similarity with the treated one.

2.6 Methodology

The evaluator used the information from the employment registry in a **difference-in-difference (DiD)** set-up. Treated (group A) and control units (group B) were observed both before and after the treatment, i.e. in the pre- and post-intervention periods, on a monthly basis. The treated group consisted of eligible individuals (i.e., long-term unemployed) in the five targeted regions of the country where the programmes were implemented. The control group was instead constructed by selecting individuals satisfying the eligibility criteria (i.e., long-term unemployed) but living in the non-targeted regions of the country where the programme was not put in place. Importantly, in order to ensure that the control units were exposed to macroeconomic conditions similar to those of the treated units, only some non-targeted regions were considered. In particular, out of the five regions where the programme was not implemented, only the three regions with a share of long-term unemployed in the working-age population equal to 8-9.9% were taken into consideration to construct the group.

Therefore, the data from the employment registry described above was used firstly to identify **potential counterfactual units** by selecting eligible units in the regions where the programme was not implemented. Secondly, it was used to check whether before the policy kicked in eligible units were similarly affected by macro trends, independently of the regions of residence.

The most relevant assumption in this setting, namely the **'parallel trends' assumption**, postulates that in the pre-intervention period the outcome variables in the treated and control groups follow similar trends. In this case, the outcome of interest was the proportion of long-term unemployed leaving unemployment, measured by the rate of outflow from unemployment to employment. The outflow was measured between the 12th and 24th month of an unemployment spell, i.e. after the one-year unemployment period, which qualifies them as long-term unemployed.

The 'parallel trends' assumption was hence verified by observing the proportion of long-term unemployed leaving unemployment for work between the 12th and 24th month of unemployment, in targeted and non-targeted regions in the 2 years before the intervention. If the 'parallel trends' is valid, the average change observed in the post-intervention period in the control group represents the **counterfactual change in the treatment group** if there was no treatment.

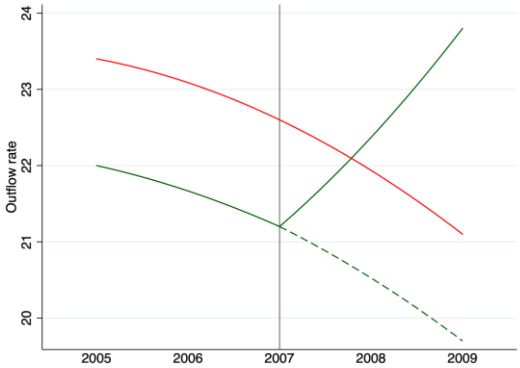
After verifying the validity of this assumption, the effect of the training programme was therefore estimated by comparing targeted and non-targeted regions. The change over time in the proportion of the long-term unemployed leaving unemployment for employment (difference between the pre- and post-treatment periods) observed in the targeted regions was compared with the change observed in non-targeted regions over the same period, to calculate the difference in the observed changes. This **double differencing** allowed identifying the change over time attributable to the training programme, as a measure of the treatment's effect.

2.7 Interpretation of results

In Figure 3 it is possible to assess the validity of the 'parallel trend' assumption by observing the evolution over time in the pre-intervention period of the outcome of interest in the targeted and non-targeted regions.

The vertical line indicates the year of the intervention's implementation, i.e. 2007. The pre-intervention period corresponds to years 2004-2006, while the post-intervention period spans from 2008, when the intervention was completed, to 2010. For the assumption to be valid, between 2004 and 2006 the outflow rate into employment between the 12th and 24th month of an unemployment spell should evolve similarly in the two groups of targeted (green line) and non-targeted (red line) regions.

Figure 3: Rate of outflow from unemployment to employment



Caption: The outflow rate (%) is calculated for each month in the time interval 2005-2009. The lines shown are the results of the interpolation of the points corresponding to the monthly values of the rate of outflow into employment.

As shown, the series of the outflow rate in the targeted and non-targeted regions followed a similar decreasing trend in the 24 months before 2007, with the distance between the red and the green line being constant over time until the introduction of the policy in 2007. **In the post intervention period**, i.e. after 2007, **the rate of outflow into employment showed a sizeable increase** in the targeted regions (green line) while it continued following the decreasing trend in the non-targeted regions (red line).

The precise impact was estimated by calculating the difference between the observed values of the outflow rate from unemployment to employment in the targeted region (green line) and the hypothetical, counterfactual values (dotted green line) in the post-intervention period. This counterfactual path, i.e. the trend that would have been observed in the targeted region had the intervention not been in place, was estimated on the basis of the starting value of the outflow rate in the targeted region observed in 2007 (21.2%), and of the trend inferred from the non-targeted region (decrease from 22.6 to 21.1% between 2007 and 2009). The comparison between the real (23.8%) and counterfactual (19.7%) values of the outflow rate for 2010 results in an estimated impact equal to 4.1%. Although small in magnitude, the results of the evaluation show that the increase is statistically significant (see previous example for the definition of statistical significance).

The increasing flow into employment of the treated group as compared to a stable decreasing flow of the control group supports the hypothesis that **the policy measure was effective.**

2.8 Lessons learnt

Advantage of DiD method

The **acquisition of all data** for both treated and control groups, **both before and after** the implementation of the intervention is critical in applying the DiD.

The DiD method exploits **the longitudinal dimension** of the data and is based on the assumption of 'parallel trends' in outcome variables of the treated and control groups in the pre-intervention period.

The double differencing (Imbens and Wooldridge, 2009) at the core of this method can be interpreted in a twofold way, namely as:

i) the difference in the post-intervention period between the outcomes of the treated and control groups, controlling for the difference between the two groups in the pre-intervention period; or
ii) the difference over time between the post- and the pre-intervention period of the outcome of the treated group, controlling for the difference over time in the outcome of the control group.

The most important advantage of the DiD method is that in the post-intervention comparison (case i), data on the pre-intervention period can be used to control for **possible pre-existing differences** between the two groups in the pre-intervention period. With this it is possible to address the potential estimation bias and properly compare outcomes in the post-intervention period, while also controlling for differences in time-constant unobservable characteristics. Analogously, in the comparison over time (case ii), information on the trend of the outcome variable observed in the control group is used to take into account possible trends unrelated to the treatment (arising both from observable and unobservable characteristics), tackling potential bias in the estimation of the treatment effect.

Gap in data

The biggest challenge faced by the evaluators arose from the fact that information on labour market history retrieved by the employment registry was not available for the entire pre-intervention period considered. Therefore, to **fill the gap in the labour market history** and reconstruct the entire path of unemployment/employment episodes of individuals before they started the training, the evaluators had to make use of information from **other data sources**. Using the social security database for this purpose was very useful to address this issue.

3. Example of the propensity score matching and coarsened exact matching methods: Internship programme for the older unemployed

3.1 Policy context

Given the increase in the unemployment rate as a consequence of the 2008 economic crisis, in 2011 the ESF managing authority of country C launched an **internship programme targeting unemployed individuals aged over 50**, representing those more severely hit by worsening labour market conditions.

3.2 Policy objectives

The internship programme could last between 3 and 6 months and was aimed at **enhancing the skills and competencies of the older unemployed** in order to facilitate their **mobility to different sectors/industries** and their active re-entry into the labour market.

3.3 Outcomes of interest

The effectiveness of the internship programme was measured by looking at the labour market outcomes of the targeted individuals, i.e. the employment status of the older unemployed.

Given the policy objectives, the main outcomes of interest were the:

1. **skills and competencies** of the treated older unemployed;
2. **labour market status** of the treated older unemployed;
3. **probability** of the treated older unemployed finding a job in a **sector / industry** different from the previous job before the unemployment spell.

The evaluation focused on the second type of outcomes.

3.4 Evaluation plan

The evaluation was planned ex ante, at the time of the programme's launch. This allowed the evaluators to easily prearrange the linkage of monitoring data from the public employment service (PES) with data from the tax records registry.

3.5 Data description

The dataset used for the evaluation was built by merging **two administrative data sources**: 1) employment data from the PES, and 2) the tax records registry.

The **PES** collected:

- **intervention monitoring data**, i.e. information on participation in the programme, such as:
    - details on the start and end dates of the internship;
    - the application date;
    - information on the firms where the internship was performed;
- **standard information on the population of the registered unemployed**, such as:
    - demographic characteristics;

- start/end dates of unemployment spells.

The **tax records registry** of the population was used to retrieve information on the work career of both people participating in the intervention and those selected as control units. While the employment status in the post-intervention period was used to assess the outcome of the intervention, the work history of individuals prior to the intervention provided a key control variable for the analysis.

3.6 Methodology

In order to evaluate whether the internship programme improved the employment prospects of the older unemployed individuals, matching techniques were applied.

The purpose of these techniques is to estimate the **treatment effect** by comparing treated units with control units that are similar in terms of relevant observable characteristics, i.e. those observable features that influence simultaneously treatment participation and outcome variables. The outcome value of each treated unit is compared with the outcome value of the matched control unit that is identical to the treated in terms of these observed characteristics.

The parameters of interest that can be estimated by means of matching procedures are the **average treatment effect (ATE)** and the **average treatment effect on the treated (ATT**), which differ in the population they refer to. While the ATE measures the average impact for a randomly selected unit from the population, the ATT refers to the subpopulation of intended recipients and measures the average impact for the units who were actually treated. In the following, ATT estimates are presented.

The main assumption of the matching procedures is the '**selection on observables**', also referred to as 'conditional independence assumption', 'unconfoundedness', 'exogeneity', or 'ignorability'. It states that given a set of observable characteristics, which are not affected by the treatment, potential outcomes are independent of treatment assignment. This implies that (i) the selection into treatment depends only on observable characteristics, and (ii) relevant characteristics can be observed. Controlling for these and adjusting for differences between the treated and control group ensures that the comparison between these two groups can lead to an unbiased estimate of the treatment effect.

The techniques used were **propensity score matching (PSM)** and **coarsened exact matching (CEM).** Their main difference lies in the way the curse of dimensionality problem is tackled. This arises from the high number of relevant characteristics to be considered to match units.

The **PSM** is based on the estimation of an indicator that summarises all information on individual characteristics. This is defined as the probability of being assigned to the treatment conditional on the observed characteristics (Rosenbaum and Rubin, 1983) and is called propensity score (PS). If the PS is correctly estimated, individuals with similar values in this index are also similar in terms of observable confounding factors and potential counterfactual outcomes (Abadie and Cattaneo, 2018). Therefore, once the PS is estimated, treated units are matched with control units having the same values of the estimated PS.

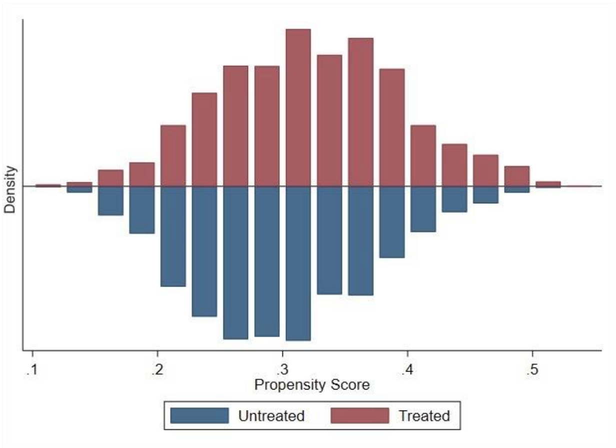
In a similar way, the **CEM** is aimed at achieving balance between the treated and control units in terms of characteristics (Iacus, King and Porro, 2008). Instead of using a single indicator for matching similar treated and control units, however, the CEM is based on a coarsening procedure. This implies that each variable is temporarily divided into intervals, so as to have an exact match on these coarsened data, identifying units who are similar with respect to classes of values of their characteristics.

These techniques were used to construct a counterfactual based on information such as the job and socio-economic characteristics relevant to depicting the programme selection process. Importantly, the work history of individuals prior to the intervention was also used as a control variable for the matching procedure. In the PSM approach, the number of months worked in the 24 months prior to the internship launch was included among the individual characteristics used to estimate the PS. In the CEM approach, the same variable was utilised to balance characteristics of treated and control units, by using the following intervals of number of months: 0-6, 7-12, 13-18, 19-24.

A selected sample of the registered unemployed from the PES registry was used to construct the **control group** for the analysis. The selection was based on the last date of registration at the PES, considering control units who registered in the same time period as that of participants.

A second assumption for the use of matching techniques is the '**common support**' or overlap condition. This assures that for each value of the PS (or for each stratum in CEM), there is a positive probability of being either treated or untreated, and therefore that there is sufficient overlap in the characteristics of participants and non-participants in the treatment to find adequate matches. For the PSM, Figure 4 on the common support proved the validity of this condition, since for all values of the PS of treated units (red bars), it is possible to find matchable control units with the same values (blue bars).

Figure 4: Common support



Caption: The 'Common support' graph shows the distribution of the estimated PS in the two groups of treated and untreated units.

3.7 Interpretation of results

The matching procedures were used to estimate the **ATT** on the outcome of interest, i.e. the probability of being employed following the internship programme. The ATT parameter was estimated by calculating the average difference between the outcome of the treated units and the outcome of the matched control units.

Table 1 shows the estimates of the impact of participating in the internship on the probability of being employed 1 year after completion of the internship. These are estimates obtained using the two matching approaches; the results from the two types of matching that were applied, namely PSM and CEM, point **towards not statistically significant results**. The value 0.038 estimated through the PSM is the result of the difference between the average employment probability of treated units, equal to 0.24, and matched untreated units, equal to 0.202. The value 0.041 resulting from the CEM corresponds to the difference between the average employment probability of treated units, equal to 0.322, and matched untreated units, equal to 0.281. The p-values for the ATT estimated both with the PSM and the CEM are higher than 0.05, therefore providing weak evidence against the null hypothesis of an ATE equal to 0.

Table 1. Estimated ATT 1 year after completion of the internship.

PSM		CEM	
ATT	p-value	ATT	p-value
0.038	0.125	0.041	0.146

Lock-in effect

As shown in other evaluations of training programmes, **participation in an internship in the intervention period might reduce the job-search efforts** of participants, hence undermining their employment prospects and inducing the so called 'lock-in' effect. This could explain the lack of effect of participation in the programme in the short term. Further analysis could be performed to evaluate the programme's effectiveness in the long term.

3.8 Lessons learnt

Alternative method

The presence of a well-defined threshold in the eligibility criteria based on age (>50) opens the possibility of applying a different quasi-experimental method, namely the **RDD**. This could be used in order **to exploit the local random assignment** around the threshold of 50, consequently **guaranteeing the comparability** between the unemployed above (treated) and below (untreated) the threshold. However, in this particular case, the evaluators stated that the RDD approach could not be applied because **the size of the control group** of registered unemployed who did not benefit from similar interventions below the threshold, i.e. in the age class 45-50, was too small.

Since the evaluation did not provide information about the exact number of control units in the age class 45-50, it is not possible to assess the validity of this methodological choice. However, it should be noted that in some RDD applications, it may be useful to **choose different bandwidths on each side of the threshold** (Cattaneo, Idrobo, and Titiunik, 2019) so as to obtain two (one-sided) estimates for the treatment and control groups, and then calculate the difference to estimate the RDD treatment effect.

Administrative data

The matching proves to be a reliable approach given the availability of administrative data that allows all relevant (observable) characteristics of the unemployed to be controlled for. This permits making the 'selection-on-observables' assumption more plausible.

Administrative registries are indeed a rich source of information, often allowing reconstruction of **the previous labour experience** of individuals. As shown in the evaluation literature, the latter is crucial as a proxy for individual skills and motivation and hence **to control for unobservable features of treated and non-treated individuals**.

Results by sector and gender

Since internships were carried out in firms in a wide range of sectors, differing also in terms of female employment, the analysis could be improved by estimating possible heterogeneous results by sector and by gender.

4. Example of the propensity score matching method: Promotion of enterprises

4.1 Policy context

In September 2014, country D's operational programme, the Innovative Economy, launched the 'Promotion of enterprises' intervention, providing **training programmes for business managers**. The intervention took place throughout the programming period 2014-2020. The training courses for managers had a **fixed duration of 6 weeks** and focused on business-oriented competence development.

4.2 Policy objectives

The aim of the training programmes was:

(i) to strengthen the managers' management skills;

and, indirectly,

(ii) to boost enterprises' productivity and growth.

4.3 Outcomes of interest

Given the policy objectives, the intended outcomes of the intervention were:

(i) **improvement in managers' skills**; and

(ii) **growth** in **sales**, **revenue**, and **productivity** of enterprises benefiting from the intervention as a result of the training courses attended by the enterprises' managers.

The evaluation focused on the second type of outcomes.

4.4 Evaluation plan

The interim evaluation was not planned ex ante but decided only at the end of 2016 in order to analyse the first 3 years of the intervention implementation. It was carried out in December 2017.

In 2015, at the end of the first year of implementation, an **ad hoc survey of the programme** was carried out to collect additional information on participating managers from the firms. Given that at this time the evaluation was not yet planned, there were no plans to use this data for evaluation purposes, and the people surveyed were not asked their permission. This generated data-access issues due to privacy requirements when the evaluation was carried out. As a result, the survey data could not be used for the evaluation.

4.5 Data description

The impact of the 'Promotion of enterprises' intervention was measured by changes in enterprises' sales, revenue, and productivity. The unit of observation was therefore represented by enterprises, and the evaluation was based on **enterprise-level data from the administrative registers of businesses** provided by the **Enterprise Agency** of country D.

The data included enterprise characteristics such as:

- location

- industry
- economic sector
- size
- sales
- revenues
- productivity (calculated as revenues divided by the number of employees).

The Innovative Economy operational programme conducted a survey to obtain the **details on managers attending the training programme**.

Interim evaluation

The data used for the interim evaluation of the first 3 years of the intervention spanned from 2014 to 2017. Data on enterprises was extracted on a monthly basis and the outcomes were measured in September 2017.

The treated group consisted of 1,120 enterprises, while the control group included 4,800 enterprises whose managers did not benefit from training courses.

4.6 Methodology

An enterprise was considered to benefit from the 'Promotion of enterprises' intervention if at least one of its managers participated in the training programme during the 2014-2020 programming period.

Since the effects of the intervention were measured on enterprise performance in terms of growth of employment, revenue, and productivity, the evaluation was consequently performed at enterprise level. **Enterprises involved in the intervention were compared with their counterparts that were not involved** by means of **PSM.**

As a first step, several enterprise characteristics (namely location, industry, economic sector and size) were used to estimate the PS. As a second step, treated and control firms were matched based on the values of the estimated PS. The intervention impact was then evaluated by estimating the ATT for each outcome variable of interest (i.e. sales, revenues, and productivity) through the difference between the average outcome in the treated group and the average outcome in the control group.

4.7 Interpretation of results

The ATT values presented for sales, revenues, and productivity are the result of the average difference in outcomes between treated units and their matched untreated units, measured in September 2017. While **no positive impact** could be observed **on the sales** of treated enterprises, the evaluation results showed a **positive change** in the **revenues and productivity** of enterprises. The ATT estimated on firms' sales was indeed equal to 12% but not statistically significant, while the ATT estimated on revenue and productivity outcomes were both statistically significant and equal to 0.8% and 0.64% respectively.

4.8 Lessons learnt

Missing information

Although one can see a positive change in the revenue and productivity of the enterprises, the estimated results should be viewed with **caution,** since some steps of the estimation are not satisfactorily explained.

Information is missing on the following aspects.

1) The **selection mechanism**

The selection mechanism takes place in two stages: (i) at enterprise level; and within enterprises, (ii) at the level of the managers who are selected. The information about enterprise selection would serve to assess whether these self-select for the treatment or if the selection is based on observable characteristics. In the second case, the key assumption of the matching approach, i.e. the 'selection-on-observables' assumption, would be valid.

2) The **enterprises selected as control units**

One would like to check if after the conclusion of the 'Promotion of enterprises' programme, firms considered as controls benefited from other types of intervention taking place in the period under scrutiny for the assessment of the outcomes.

3) Heterogeneity of the results

Since firms belong to different sectors and are heterogeneous in size, one could also assess if the estimated training impact is heterogeneous across different types of firms. Moreover, additional information could be used to investigate what are the factors influencing the estimated results. Possible drivers to be controlled for could be: (i) number of managers participating in the training programme within each firm; and (ii) managers' characteristics (tenure, field, level of an education, and previous work history), to proxy for unobservable factors such as managerial skills and motivation.

Planning evaluation ex ante

The additional information gathered through the ad hoc survey of the programme for monitoring was not made available for evaluation. Planning the evaluation in advance would have been useful so that plans could be made to use the survey data for evaluation purposes, preventing data-access issues, and lowering the cost of the evaluation.

References

- Abadie, A., and Cattaneo, M. D. (2018). Econometric methods for program evaluation. Annual Review of Economics, 10, 465-503.
- Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2019). A practical introduction to regression discontinuity designs. Series: Elements in Quantitative and Computational Methods for Social Science, Cambridge University Press.
- Iacus, S. M., King, G., and Porro, G. (2008). Matching for causal inference without balance checking. http://gking.harvard.edu/files/cem.pdf.
- Imbens, G. W., and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. Journal of Economic Literature, 47 (1): 5-86.
- Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Further reading on CIE

- Angrist, J.D., and Pischke, J. (2008), Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press.
- European Commission (2013): Design and commissioning of counterfactual impact evaluations. A practical guidance for ESF managing authorities. Publications Office of the European Union.
- European Commission (2019): Advanced counterfactual evaluation methods - Guidance document. Publications Office of the European Union.
- European Commission (2020): How to use administrative data in ESF evaluation - Guidance document. Publications Office of the European Union.

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct Information Centres. You can find the address of the centre nearest you at: http://europa.eu/contact

On the phone or by e-mail

Europe Direct is a service that answers your questions about the European Union. You can contact this service

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by electronic mail via: http://europa.eu/contact

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU Publications

You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/contact)

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en/data) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

Publications Office
of the European Union