

COUNTERFACTUAL EVALUATION OF YOUTH EMPLOYMENT POLICIES

METHODOLOGICAL GUIDE

MÁRTON CSILLAG, JUDIT KREKÓ, ÁGOTA SCHARLE

With contributions from:

MARCO CENTRA, MASSIMILIANO DEIDDA, LUCÍA GORJÓN,
VALENTINA GUALTIERI, IGA MAGDA, MARTA PALCZYŃSKA,
FRANCESCO TRENTINI, CLAUDIA VILLOSIO

December 2020



Counterfactual evaluation of youth employment policies

Methodological guide

The “Youth Employment PartnerSHIP” project is funded by
Iceland, Lichtenstein and Norway
through the EEA and Norway Grants Fund for Youth Employment.

IMPLEMENTED BY:



CONTENTS

1. INTRODUCTION	4
2. ADMINISTRATIVE DATA	8
3. THE PROCESS OF THE EVALUATION	13
3.1. Choice of the specific programme	13
3.2. Choice of the outcome variables	13
3.3. Identification strategy: choice of the counterfactual impact evaluation (CIE) method	15
3.4. Heterogeneous effects	21
4. ISSUES RELEVANT FOR POLICYMAKERS	23
4.1. Interpretation of the results	23
4.2. Link the results to programme design and implementation	24
4.3. External validity, comparability with other studies, and aggregate statistics	24
4.4. Displacement effect and deadweight loss	24
4.5. Cost-benefit and cost-effectiveness analysis	25
References	26
Appendix 1: Short description of the evaluations of the four countries	28
Italy	28
Hungary	29
Poland (Evaluation 1)	29
Poland (Evaluation 2)	30
Spain	31
Appendix 2: Glossary and abbreviations	32

1. INTRODUCTION¹

This guide provides a step-by-step introduction to the counterfactual evaluation of labour market policies for youth with a focus on the use of administrative data. The main issues are illustrated by the practical problems encountered in evaluating hiring subsidies for youth in the four countries of the Youth Employment PartnerSHIP project (Spain, Hungary, Italy, and Poland). However, the guide can be applied to evaluating other programmes as well.

The purpose of this guide is to provide a practical toolkit for researchers and managers of institutions carrying out evaluations (research, ministries, public employment services) who plan to initiate and coordinate youth employment programme evaluations. The guide assumes a basic understanding of statistical and evaluation concepts, and provides a glossary of key technical terms. As the content and the aim of the programmes differ across countries, and the available data sources that can be used for the evaluations are also diverse, this guide does not present a unified evaluation framework.² Instead, it summarises and illustrates common practical issues and problems that are specific to the counterfactual evaluation of policies designed for youth employment.³

Though Youth Guarantee programmes⁴ were introduced in all EU member states, the programme elements have not been rigorously evaluated in all of the states. Policy evaluation is crucial for understanding the impact of the policies, and data-driven quantitative evaluations are key if we want to know whether the goals of the policies have been reached, and for improving the efficiency of the programmes. Administrative data may allow for ex-post evaluations even in countries where the programmes were introduced without an extensive monitoring and data collection framework.

The aim of counterfactual evaluations of labour market policies is to estimate the causal effect of a programme on different outcomes, such as the probability and duration of being employed. This implies an estimation of the effects of the policy compared to the counterfactual state in which the policy was not introduced, all else being equal. In practice, as the counterfactual world cannot be observed, the key challenge is to find a credible control group made up of individuals who can be regarded as similar in all respects to the programme participants, except that they do not take part in the programme. At all stages of the evaluation, theoretical considerations and constraints stemming from data limitations also play a role in the researchers' choices.

The Youth Employment PartnerSHIP project evaluates labour market policies that primarily act on the demand side by providing monetary incentives to employers for hiring young people. In two countries (Hungary and Poland), the project focuses on the hiring subsidies offered in the Youth Guarantee, which is a commitment of all EU member states to provide young NEETs with a good quality offer of employment, education, apprenticeship, or traineeship within a short

1 We are grateful to Namita Datta, Ilf Bencheikh, Tibor Keresztély, and András Svraka for their valuable comments.

2 This guide concentrates on counterfactual evaluations implemented on administrative datasets. For more comprehensive monitoring and evaluation guides, starting from the development of a theory of change, see, for example, Leeuw, & Vaessen(2009) and UNDP(2009), Gertler et al. (2016.), Morris et al. (2013).

3 For detailed descriptions of counterfactual quantitative evaluation methods, see, for example, Angrist and Pischke (2009), Abadie and Cattaneo (2018), Gertler et al. (2016) and Morris et al. (2013), European Commission (2013).

4 These are financially supported by the Youth Employment Initiative (YEI) in regions where the youth unemployment rate was higher than 25%.

period of time. In two other countries (Italy and Spain), the evaluated measures aim to increase the stability of the employment of young people by incentivising hiring with permanent (open-ended) contracts. The main elements of the evaluations are sketched in Boxes 1 and 2, and are summarised in more detail Appendix 1.

All four evaluations of the Youth Employment PartnerSHIP project were carried out using some kind of individual-level administrative data (unemployment registers, social security records, compulsory notices records of the flows in and out of formal employment, etc.), which are available in most European countries. This guide focuses in particular on the challenges and the problems that arise when using administrative data in policy evaluations.

The guide is structured as follows. The first part summarises the most important issues related to accessing and using administrative data, since data constraints have an impact on the evaluation process at all stages. The second part outlines the process of quantitative evaluation, concentrating on the choices researchers must make to ensure that they obtain valid results. The final part highlights the issues and the problems that can emerge in the subsequent stages related to questions that arise when presenting results to policymakers. The sections are based on and illustrated by the experiences and specific examples of the four countries of the Youth Employment PartnerSHIP project. The main elements of the evaluations are sketched in Boxes 1 and 2, and are summarised in in more detail Appendix 1. Appendix 2 provides definitions of the key technical terms.

Box 1: Details of the country evaluation: Italy

ITALY	
Policy	Jointly evaluated: 1) Lowered firing costs for open-ended contracts (Graded Security Contract) 2) Social security rebate to new open-ended contracts and to conversions from a fixed-term to an open-ended position)
Eligibility criteria	Persons who had not had an open-ended contract in the six months preceding their hiring.
Duration, type	1) Permanently changes the regulation on dismissal. 2) 36 months of hiring subsidy/incentive
Subsidy	1) Reshape of regulation of dismissals 2) 100% rebate of non-wage labour costs
Data	Sisco (statistic system of online mandatory communication). Sisco is the public administrative registry with elementary information on hires, conversions, terminations.
Sample	Employment contracts registered in 2014-2015.
Methodology	DiD, with parametric correction for sample selection and interactions between the treatment and the different age classes, in order to retrieve the effects on young cohorts
Outcome	Share of new hires with an open-ended contract over the total employment contracts registered in 2015

Box 2: Details of the country evaluation: Poland

POLAND	
Policy	1) Wage subsidy 2) On-the-job training, classroom training, public works, wage subsidy (intervention works), on-the-job training voucher, and classroom training voucher
Eligibility criteria	Unemployed under 30 years old for both (1) and (2)
Duration, type	2) Employer-side subsidy up to minimum wage + social security (~€500 + €100 in 2018) (2) subsidy to public and intervention works and on-the-job training vouchers
Subsidy	2) Employer-side subsidy up to minimum wage + social security (~€500 + €100 in 2018) (2) subsidy to public and intervention works and on-the-job training vouchers
Data	PES registers
Sample	2015–2017
Methodology	1) RDD combined with DiD 2) propensity score matching
Outcome	1) Being off the unemployment register 12 to 36 months after the initial registration and 2) Being off the register and not on any ALMP 12 to 36 months after registration

Box 3: Details of the country evaluation: Spain

SPAIN	
Policy	Internship contract, IC
Eligibility criteria	Bachelor's or vocational training degree, under 30 years old, unemployed
Duration, type	minimum six months, maximum two years. Full- or part-time
Subsidy	50-75% reduction in SSC contribution. Bonuses for the conversion to permanent contract of €500 per year for a man or €700 for a woman, for the first three years.
Data	(1) PES registers, complete register of contracts signed, linked to (2) A sample of social security records: continuous sample of work histories (CSWH)
Sample	2002–2018
Methodology	Multinomial logit and logit model approach to compare internship participants with eligible non-participants under regular temporary contracts.
Outcome	1) Probability that the individual remains at the firm after the IC. 2) Probability that the individual signs a permanent contract at the same firm. 3) Probability that the individual signs a permanent contract at another firm.

Box 4: Details of the country evaluation: Hungary

HUNGARY	
Policy	90-day job trial wage subsidy
Eligibility criteria	Under 25 years old, registered unemployed
Duration, type	Maximum 90 days
Subsidy	100% of wage costs (wage + employer's contribution)
Data	1) PES registers linked to 2) Social security records and education records
Sample	2015–2017 (history variables are constructed beginning from 2003)
Methodology	1) Nearest-neighbour and propensity score matching to compare job trial participants with public works and classroom training participants 2) Difference-in-differences, exploiting that the programme started 9 months later in one region
Outcome	1) In work 6 and 12 months after start of subsidy 2) Cumulative wages in 6 and 12 months after start of subsidy

2. ADMINISTRATIVE DATA

Administrative data are usually characterised by high or full coverage and large sample sizes, and in many cases there is an option to link to different data sources.⁵ While the reliability of administrative data is usually higher than that of survey data, administrative data can be also incomplete or uncertain in many cases. In labour market policy evaluations, the most common administrative datasets are social security data, unemployment registry data, public employment services (PES) data (including the details of labour market programmes and their participants), and employer databases. As administrative datasets contain individual- or firm-level data, access to administrative datasets is usually strictly regulated and restricted based on data protection considerations. Both the way administrative data are collected and stored and the legal environment in which datasets are obtained and handled differ across countries. Hence, the evaluator might face difficulties and a long administrative process in accessing the required data. Many issues arise when using administrative data⁶:

- **Does a country-level administrative database exist?**

The scope of the evaluation is influenced by whether the administrative data in question are registered and processed in a country-level database, or at the level of several administrative territorial divisions. In **Hungary, Spain, and Poland**, the PES data are integrated into a country-level database, while **Italy** has a country-level administrative register of work relationships named SISCO.

- **Is there any scope for linking the dataset to other administrative data in order to enrich the set of control and outcome variables?**

Generally, different forms of information on a given person are registered in various databases managed by different authorities, such as social security institutions or employment offices, training providers, tax authorities. Linking these different sources could greatly increase the scope of the available information. For example, in **Hungary**, PES data can be linked to administrative data of the pension authority that contain data on the employment and the wages of all registered jobseekers who show up in the PES data. This allows the researcher to use more outcome variables (such as the quality of employment), as well as a richer pool of observable characteristics (including a history of prior employment spells). In **Spain**, the FEDEA has access to two administrative databases that can be linked to each other: the sample of social security records (CSWH) and the PES unemployment registers and the contracts signed monthly in Spain. However, in many cases, the linking of different data sources is strictly regulated and limited or prohibited for data and privacy protection considerations. For example, in **Poland**, PES data cannot be linked to the data of the social insurance institution on employment and wages, except in cases precisely defined by law, such as reporting to the European Commission.

⁵ See, for example, Caliendo et al. (2017) on the added benefits of having access to survey-based data for the evaluation of ALMPs.

⁶ For a step-by-step guide for the use of administrative data for European Social Funds counterfactual impact evaluations, see also EC(2020).

- **What is the procedure for accessing the data?**

As there is no EU-wide regulation regarding access to these data, the administrative requirements for obtaining access differ across the countries. For example, in Spain, access to the database of working histories (CSWH) can be obtained by submitting a request to the Spanish social security agency. The PES data are generally not readily available to researchers, but the FEDEA was granted access after a cooperation agreement was signed by the two parties. However, there are no formal procedures for obtaining access to social security and PES data in Poland, where the process is discretionary.⁷

- **Institutional background and administrative use of the data**

Researchers are strongly advised to contact the data owners in order to clarify the institutional background, the original purpose, and the processes and routines of data collection. The use, and thus the reliability of the quality of the data might differ even within a single database, depending on whether a given piece of information is used in any accounting, evaluating, or administrative procedures. A deep knowledge about the mechanisms of the data collection, storage, and processing is essential in setting up an estimation strategy.

In some cases, especially depending on whether the data collection and the evaluation fall under the authority of a single or several cooperating organisations, the evaluator may seek to influence the method and the content of the data collection so that they better serve the specific purpose of the evaluation. The feedback loop between the process of data evaluation and data collection is especially useful for long-term or regular evaluation tasks.



⁷ Poland is currently building an integrated analytical platform that will be a centralised system, which will ease the analysis of administrative data from various sources.

- **Quality and completeness of the data**

As administrative datasets typically cover the complete set of observations, or a random sample from all observations, the representativeness of the data is usually ensured. However, administrative datasets may also contain serious errors and missing data, especially in the case of variables for which the filling is optional. The researcher must always assess the magnitude and the randomness of the missing data and apply corrections if necessary. The general completeness and the quality of the dataset can, for example, be checked by comparing it with aggregate data from alternative sources.

- **Are the same data available for the treatment and the control groups?**

A frequent problem that can arise in policy evaluations using PES data is that the database (or the linked database) may contain different data (different structure, different variables from different sources) for the programme participants and for the potential control group, which could make defining common outcome and control variables more difficult. In many cases, unemployment registers contain only limited information on the real employment status of the registered jobseekers after they leave the registry or complete the labour market programme.

In **Hungary**, the PES administrative data contain data on employment status only for programme participants, and only for 180 days after completion of the programme. However, no employment data are available for registered jobseekers who did not participate in an ALMP. A possible solution to this lack of data is to use the registered unemployed status, which is also available for non-participant jobseekers, as an outcome variable. A researcher can identify whether an individual is still a registered jobseeker, and can assume that a person who has left the register is employed. However, the researcher should keep in mind that individuals can leave the unemployment register for many reasons other than having been re-employed, as the pool of people who have left the register includes many inactive individuals. In the Hungarian evaluation, this issue was solved by using the employment data of the linked social security database as an outcome variable. The situation with the **Polish** PES data is similar: i.e., an individual's employment status can only be approximated by looking at whether the person is in the register and not participating in any active programme, or by using limited information on the reason why the person is no longer registered with the office.

- **Can the programme participants be identified in the data, or only the eligible group?**

In many cases, the database does not contain information that would allow the researcher to identify the programme participants, instead, the database may provide information only on those individuals who are eligible to participate in the programme. In this case, the researcher can estimate the effect of the programme on the eligible subpopulation only, and cannot assess the effect of the programme on the participants. Using programme evaluation terminology, the researcher is able to identify the *intention to treat* (ITT), but not the *average treatment effect* (ATE) or the *average treatment effect on the treated* (ATT).

This is less of a problem if the programme is close to universal (e.g., a universal or quasi-universal voucher that all unemployed youth are entitled to, or that is mandatory for some groups). However, the distinction might be more important if participation is only partial, or if the criteria for

participating in the programme are unclear. For example, if the estimated effect of the programme on the whole eligible population is low, the researcher may be unable to tell whether the low take-up rate or the low programme effect is responsible for the weak results. This distinction matters, as the policy consequences will differ depending on whether the participation rate is low or the effect on the programme participants is low.⁸ The former might call for an intensification of outreach activities, while the latter might arise from bad programme design.

PES datasets usually contain details on active labour market policies, and allow the researcher to identify programme participants.⁹ In **Hungary**, the PES database contains details of the labour market policy programme, including the specific type, the start date, and the termination date of the programme. In the case of wage subsidies, the amount of the subsidy and of the wages are not included in the database, but these data are available from the linked social security dataset.

In **Spain**, both databases allow for the perfect identification of the internship contract (IC) under analysis. In both databases used in the Spanish evaluation, the Continuous Sample of Work Histories (CSWH) and the PES unemployment registers, all of the individuals who have signed an IC are perfectly identified, together with the start and the end dates of the contract. Furthermore, as the databases provide information on all other contract types, as well as on the internship contract, we can use them as controls. However, in **Italy**, the administrative register (SISCO) does not contain information on the programme participants. Thus, in Italy, the evaluation model estimates only the effects on the eligible population. In **Poland**, the PES database contains details of the labour market policy programmes, including information on the specific types, and on the start and the termination dates of the programmes.

- **How detailed are the set of available observable characteristics?**

The pool of observable characteristics available for use as covariates influence both the identification strategy and the scope for analysing heterogeneous effects over different subpopulations. Consequently, before building up the evaluation strategy, it is crucial to map the available or obtainable covariate variables. In individual administrative registers, the gender, the age, and some regional variables are usually available. However, beyond this very basic information, administrative databases differ in terms of the covariates they provide.

The level of education is generally seen as a crucial variable that has a large effect on employment outcomes. Thus, the level of education is among the most important variables in the evaluation of labour market policies. Information on the level of education is available in the **Hungarian** data, in the **Polish** PES data, and in the **Italian** SISCO. However, in **Spain**, information on educational attainment comes from the census, and is only updated every 10 years.

Additional useful variables are information on households (for example, number of cohabitants in **Spain**), information on different benefits (for example, maternity benefits in **Hungary**), and information on the presence of a young child in the household (Poland).

8 This issue can be alleviated by using external information to approximate the take-up rate.

9 However, eligibility typically has to be approximated based on data about the jobseekers' background characteristics.

In the case of employment records, the SISCO in **Italy** contains information on industry (NACE) and occupation. However, information on occupation is missing in the employment records in the **Spanish** social security database.

In addition to these observable characteristics of individuals, contracts, and firms, it might be useful to include in the evaluation external macroeconomic variables to control for the labour market environment. For example, in **Italy**, value added per economic sector following the hire may be used to discount the employers' expectations regarding growth at the time they decided to hire.

Moreover, the length of the time span visible in the database before the programme began might be relevant. In some cases, it might be useful to include data from before the introduction or the major restructuring of a programme in the evaluation, so that the situation prior to the policy change can be compared to the situation afterwards in a difference-in-differences framework. In other cases, variables from long data series can be used as controls. For example, in **Hungary**, the full history of past employment and registry records for each person is available from the social security database. Thus, using these records, the researcher may be able to construct the employment history of an individual in order to capture her experiences and motivations.

3. THE PROCESS OF THE EVALUATION

3.1. Choice of the specific programme

Many labour market programmes consist of multiple programme elements. The evaluator (together with the implementers) must decide which element to choose as the subject of the evaluation. The evaluation can encompass the whole package containing several subprogrammes, or it can focus on a single, well-defined, homogenous programme. This choice is constrained by the programme rules and the implementation approach. For example, if the programme elements are usually combined to meet the specific needs of each jobseeker, it may not be possible to evaluate a separate element (unless rich data on jobseeker needs are available). For example, in the case of the **Hungarian** evaluation, the 90-day job trial is often followed by a longer-term partial wage subsidy (maximum of eight months), as the programme scheme allows this combination in specific cases. Neglecting this combination would cause a serious upward bias in the employment rate six months after completion of the job trial programme, as for those individuals who have access to this subsequent subsidy, the employment rate will be close to 100%. However, most of the labour market programmes offered by the PES in **Poland** consist of one element only, such as a training programme, an internship, or an entrepreneurial subsidy. The different elements are combined only very rarely (although job search assistance is available to everyone).

The choice of the programme influences the choice of the treatment and the possible control groups. Moreover, if the participants are not identified in the data and the researcher observes only the eligible group, the evaluation of a specific programme element is not an option.

It is crucial to explore and clarify the whole design and all of the details of the programme in order to assess the scope for a potential evaluation. In this context, the following questions arise: What problem is the programme trying to solve, and how did the designers of the programme know that these were problems? Is the programme mandatory or voluntary, what are its eligibility criteria, and what is its geographical and time scope? What do we know about the targeting, sequencing, and scale of the programme? What other details are important to clarify in order to discern all of the impacts and the weaknesses of the programme? What are the typical programme combinations, and what is the consequence of these combinations on the evaluation/interpretation?

For example, the policies under scrutiny in the **Italian** evaluation are hiring incentives in the form of social security rebates, and the substitution of a permanent contract with a graded-security contract that results in a reduction in firing costs. The former policy was applied only to voluntary applicants in a fixed time window, while the second policy modified the law and applied to all new open-ended contracts after the legal change was implemented. As these two policies became effective around the same time, the evaluation estimates the combined effects of the two policies jointly.

3.2. Choice of the outcome variables

The starting point for the choice of the outcome variable and the method is to establish the impact mechanism and the goal of the programme. What are the expected effects? What is the timeline of the impacts? Are there any side effects? After we have a clear idea about the effects of the programme, the following questions arise:

- What is the outcome of interest: e.g., employment status, employment at the same firm, employment on the primary labour market, or (cumulative) wages?
- What time horizon is being analysed: the short or the long term (more than one year after the programme)? Ideally, the researcher will analyse the outcomes over different time horizons, and evaluate how the impact changes depending on the length of time. In evaluating wage subsidy programmes, the researcher should take into account that wage subsidies may be attached to an obligation of subsequent (non-subsidised) employment. Thus, the researcher should consider the outcome only after this period of obligation.

Conceptual considerations might be heavily constrained by data availability. In many cases, longer-term outcomes are not available. Moreover, the PES datasets usually do not contain wage data.

As the aim of most wage subsidy programmes is to increase the probability of an unemployed person finding a (good) job, the most relevant outcomes are the probability of the person being employed a certain time after completing the programme, and, perhaps, the person's total labour income. For example, the aim of the 90-day job trial programme that is the subject of the **Hungarian** evaluation is to introduce the unemployed person to the world of work, and to decrease the risk to the employer associated with hiring a young person without any experience. Ideally, participation in the programme would improve the person's employment prospects. Specifically, it is expected to increase the probability of the person being employed after completing the programme, and, ideally, to increase the person's wages. Thus, it is anticipated that participation in the programme will improve a young person's prospects of finding a better paid job by providing her with experience and increasing the value of her CV. Therefore, the main outcome variables are the person's employment 0.5, one, two, etc., years after completing the programme, and the cumulative wage differential over the same time horizon. However, data availability constrains the horizon of the outcome variables. In this specific case, long-term outcomes are not available, as the policy was introduced in 2015, and the available sample includes only three years; i.e., the 2015-2017 period.

Using PES data to evaluate active labour market policies (ALMP) in **Poland** gives researchers a limited choice of outcome variables. We can observe whether a person re-joined the unemployment register after x months. But the individuals who are not in the register may be working, or they may be inactive or programme participants. The outcome can be narrowed to those individuals who are not in the register and are not participating in any ALMP.¹⁰

The ultimate goals of the **Spanish** internship contract and the **Italian** hiring incentives are to increase the stability of employment and the share and the duration of permanent contracts. Therefore, the relevant outcome variables indicate whether the employment position of the young person stabilised after participating in the programme. The main outcome variables of the **Spanish** evaluation are the probability of staying at the same firm after the termination of the internship contract. The second outcome is the probability of staying at the same firm under a permanent contract. The third outcome is the probability of securing a permanent contract at a different firm after the termination of the internship contract, which also indicates that the person's labour market position stabilised as an indirect effect of the internship. The aims of the social contribution rebates and the reduced firing costs that are the subject of the **Italian** evaluation are similar: i.e., to increase the

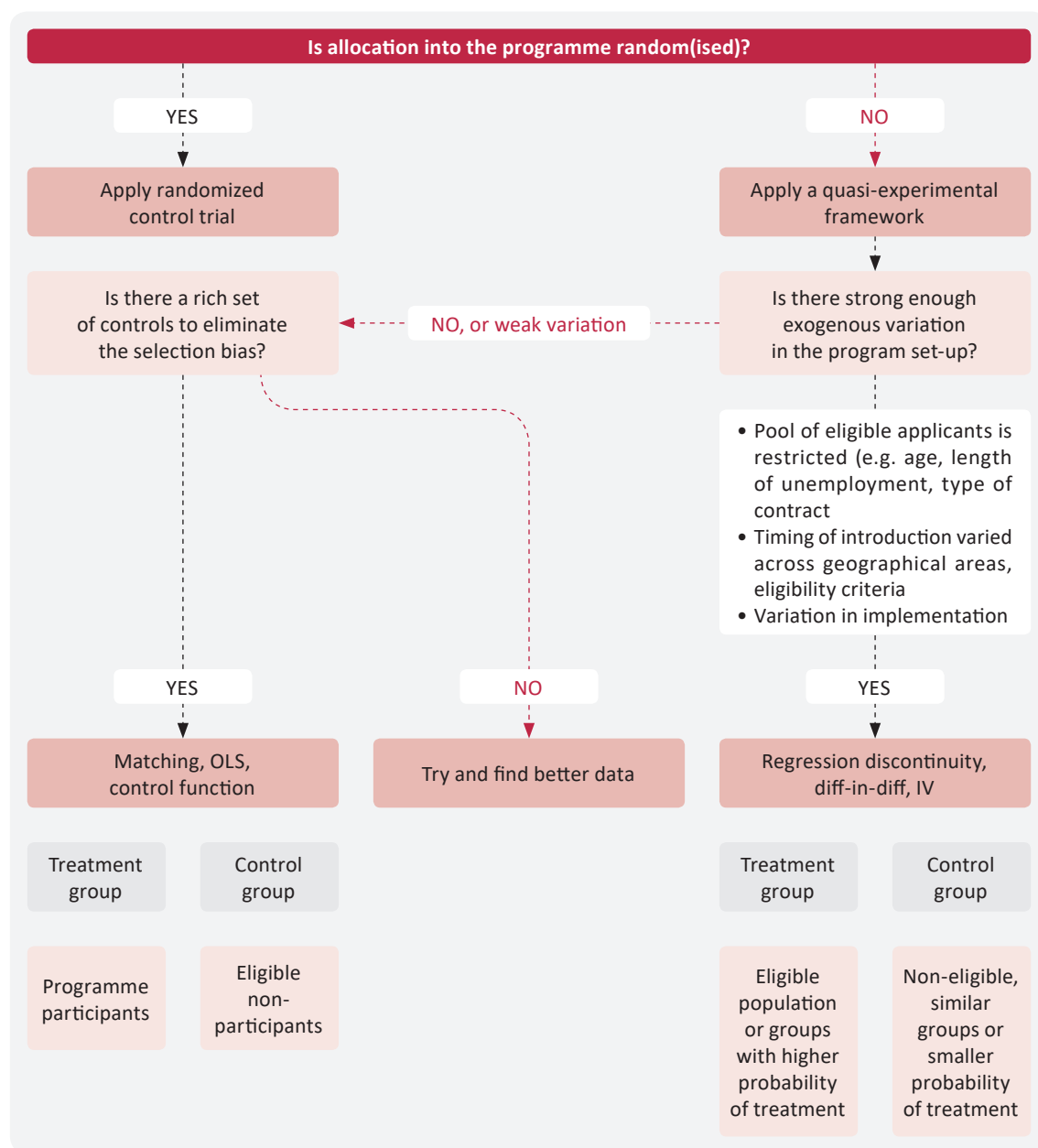
¹⁰ The time horizon of the outcome evaluated is also important. For example, the entrepreneurial subsidy in Poland requires the business to continue to operate for at least a year, or, alternatively, to repay the subsidy. Thus, its real effectiveness can be evaluated only after a year.

stability of employment and increase the duration of contracts. Thus, the main outcome variable of the evaluation of the Italian incentives for hiring with open-ended contracts is the share of new permanent contracts in 2015 over the total employment contracts registered during the year.

3.3. Identification strategy: choice of the counterfactual impact evaluation (CIE) method¹¹

The main difficulty that arises in measuring the causal impact of any policy is that the counterfactual world in which the same person does not receive the treatment is not observed. The researcher must infer the counterfactual states from observational data.

Figure 3-1: Choice of identification strategy depending on programme design and implementation



11 The basics of the most important concepts and methods are described in Appendix 2.

Ideally, the evaluation of labour market policies should be based on creating an experimental framework by randomising selection into the programme from the pool of eligible persons. When the experimental framework is not ensured, selection into the programme is not random, and non-participants might be different from participants in terms of their observable and unobservable characteristics. This might lead to two basic sources of bias. First, the outcomes, such as the employment and the wage prospects of the participants, might be different from those of the non-participants even without the programme (baseline difference). Second, the programme might affect the participants differently than it would the non-participants if they had taken part in the programme (heterogeneous treatment effect). Moreover, a third additional source of bias arises when the introduction of the treatment affects the potential outcomes of the control group as well as those of the treated group; for example, when peer effects influence the outcomes even for non-participants. (violation of the Stable Unit Treatment Values Assumption, SUTVA). In the absence of experimental circumstances, the researcher must apply a quasi-experimental framework to imitate experimental circumstances by addressing the non-random selection into the programme (see Figure 3-1).

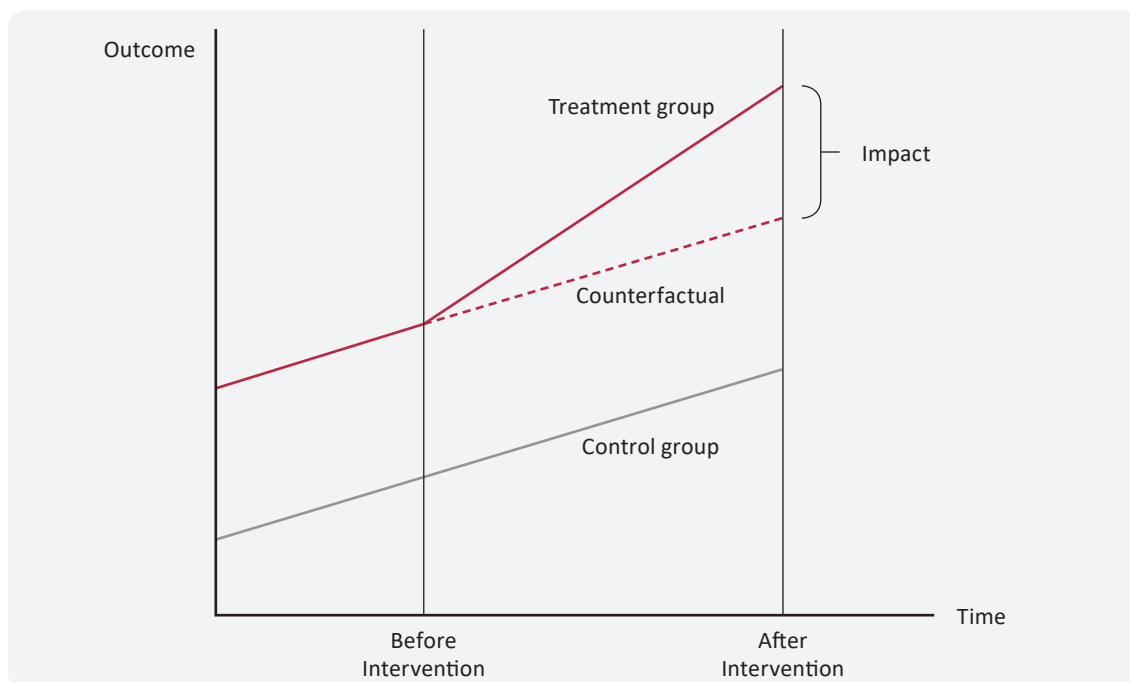
There are two basic directions for ensuring a quasi-experimental framework.

One option is to find some **exogenous variation in the programme framework**, and to exploit this exogeneity. The goal of this exercise is to distinguish between individuals who are exogenously excluded from the programme based on some eligibility restrictions defined in the programme rule, and individuals who are exogenously chosen to participate in the programme or are eligible to participate in the programme. The most common examples of exogenous restrictions that might indicate the effect of the policy are the following:

- *Pool of eligible applicants (e.g., age cut-off, cut-off in the time spent in unemployment or without a long-term contract)*

The **Italian** identification strategy exploits the fact that the hiring incentives were available only for those individuals who were not employed with an open-ended contract that had expired within the last six months. The identification strategy is based on a difference-in-differences model that compares the contracts registered in the register of employment records (SISCO) between 1 January 2015 and 31 December 2015 with those registered in the previous years. Two different groups (eligible and control) were defined in order to estimate the added share of the new open-ended employment contracts that would not have been signed in the absence of the two policies by those registered during 2015. The two groups are defined by means of the eligibility criterion. The eligible group includes people who had been hired with an employment contract during 2015, and who had not been employed with an open-ended contract that had expired within the previous six months. The control group includes people who had been hired with an employment contract during 2015, and who had been fired within six months before the beginning of the new open-ended contract while employed with an open-ended contract, and were therefore not eligible for social security rebates (for an illustration of the difference-in-difference model, see Figure 3-2).

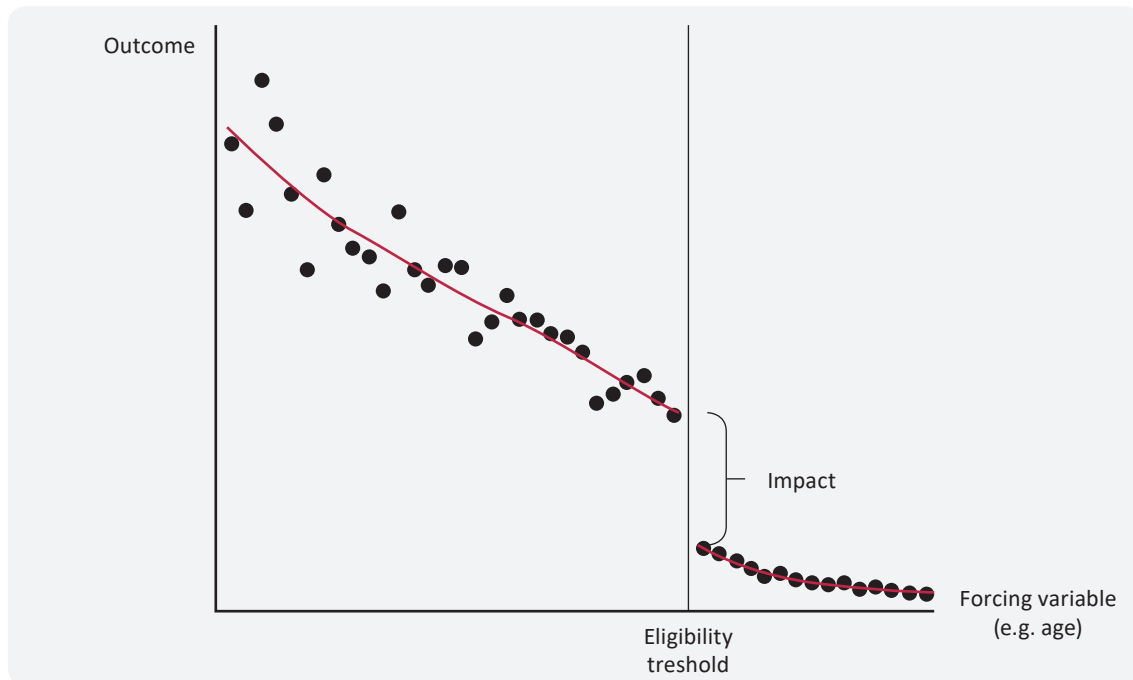
Figure 3-2: Illustration: difference-in-differences (DiD) method



As the database does not report direct information on the subsidies, and allows us to identify only the eligible individuals, the evaluation estimates the effect of the policy on the eligible group (intention to treat, ITT effect).

As Youth Guarantee programmes are aimed at young people under a certain age, most of the programmes have an age cut-off. A straightforward approach to analysing such programmes is to exploit the age cut-off in a regression discontinuity framework, as is done in one of the **Polish** evaluations. The 12-month wage subsidy programme in Poland was available for unemployed individuals under age 30. This allows us to use the age cut-off in the sharp regression discontinuity design (RDD hereafter) framework (see Figure 3 2). However, there are other institutional rules that change at the age 30 threshold. First, unemployed workers under age 30 have access to a wider range of programmes (training vouchers, reallocation vouchers, and job trial vouchers), and to ALMPs under different rules (e.g., job trials can last up to 12 months for younger workers, and up to 6 months for older workers). Second, the sources of funding partially differ between younger and older unemployed. For younger workers, funding is distributed via an operational programme that is centrally managed, while for older workers, funding is distributed via regional programmes managed at the level of voivodeships. This may lead to differences in the implementation and the availability of the programmes for unemployed workers depending on whether they are above or below age 30. The identification strategy uses the fact that before the introduction of the wage subsidy programme, all of the institutional differences between the treatment group and the control group already existed except for the wage subsidy programme under evaluation. Thus, the regression discontinuity design (RDD) and difference-in-differences (DiD) framework can be combined. An important caveat of this approach is that the results are based on a comparison only of the treatment and the control groups who are close to the cut-off age. This may limit the external validity of the results, and considerable caution is advised when extrapolating the results to other age groups.

Figure 3-3: Illustration: regression discontinuity design (RRD)

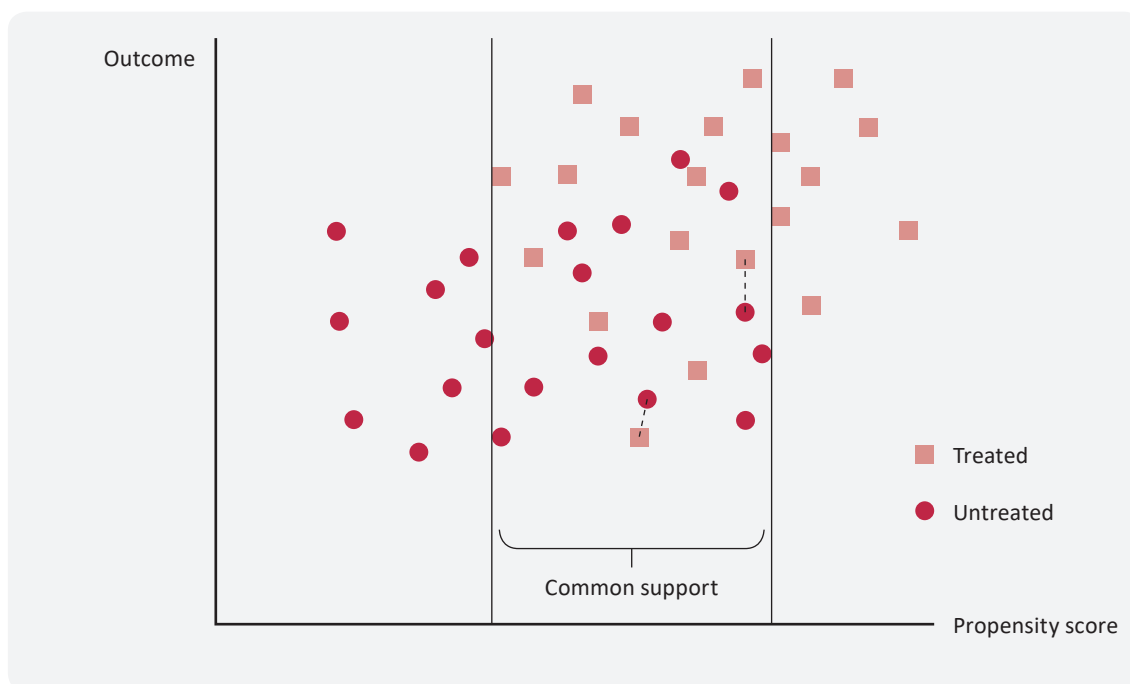


- *Regional heterogeneity in the programme intensity, start date, target numbers, etc.*

In **Hungary**, the Youth Guarantee Programme in the Central Hungarian region started in October 2015, or nine months later than in the other regions. The reason for this delay is that unlike the other six Hungarian regions, the Central Hungarian region is not a convergence region according to the EU Structural Fund convergence objective. Thus, the financial sources of the programme differed somewhat between this region and the rest of Hungary. The lag occurred for administrative reasons, and can be regarded as exogenous. The programme that was started in Central Hungary had elements and eligibility requirements identical to those of the programmes implemented in the other regions. The exogenous heterogeneity in the start date allows us to apply a difference-in-differences method as follows. The treatment group consists of eligible jobseekers in the convergence regions, and the control group consists of jobseekers in the Central Hungarian region. The treatment and the control groups are compared nine months before and after the programme started in the convergence regions: that is, in January 2015. An obvious limitation of this strategy is that by exploiting the phased implementation, we are basically comparing the treatment and the control groups for a short period; in this case, for nine months.

Other tricks that show up in the literature include exogenous “house rules” or rigour of the examiners. There are situations in which differences in eligibility or in the probability of being treated are based on differences in the informal house rules across PES agencies, counties, examiners, etc. (Maestas Mueller, 2013). Differences in the probability of being treated based on the house rules can be exploited in an instrumental variable framework.

Figure 3-4: Illustration: the propensity score matching



The other main direction in identification is to **capture selection bias with a rich set of observables**, employment history, and personal characteristics. The selection bias can be reduced by using a set of variables reflecting the personal characteristics of both the programme participants and the eligible non-participants, and, thus, by forming a group of non-participants who resemble the participants as closely as possible. In this case, the identification strategy is based on the assumption that conditional on the observed covariates, unconfoundedness is ensured; that is, adjusting for differences in the observed variables removes biases from comparisons between treated and control units. However, the main concern is that certain unobservable characteristics – above all motivation and ability – correlate with the outcome variables. For example, while a more motivated young person is more likely to apply for an active labour market policy programme, her employment prospects are also likely to be better even in the absence of the labour market policy programme. A similar argument might hold for other skills and abilities that are not captured by level of education. The researcher may try to find variables that capture unobserved abilities and motivation, such as higher competence test scores or a history of employment or participation in training programmes, that might indicate a higher level of motivation.

However, as a basic rule, these unobserved factors result in an upward selection bias of the treatment effect. Therefore, the estimated effect should be regarded as an upper bound.

In **Hungary**, the outcomes of participants of the 90-day job trial programme are evaluated by comparing them with the outcomes of participants of other programmes: i.e., training and public works programmes. While the bias arising from unobserved characteristics is assumed to be lower if the control group consists of participants of other programmes instead of non-participants, the external validity of the results is also lower. The results show the effectiveness of the programme compared to that of the other programmes. However, no conclusion can be drawn about the effect of the programme compared to a situation in which the jobseeker has not participated in any programme since the PES offices has been trying to enrol registered jobseekers

into any programme. This implies that the pool of young people who have not participated in any programmes, but who stayed in the registry, are probably not a valid control group, as they are among those with either the lowest motivation levels or the highest re-employment chances. Even the participants in the public works programme are expected to differ significantly from the job trial participants in terms of their underlying labour market prospects and motivation levels. The information in the database on each person's level of education and complete employment history allows us to mitigate the unobserved heterogeneity. The second Polish evaluation uses the same strategy to evaluate the relative effectiveness of on-the-job training, on-the-job training vouchers, classroom training, classroom training vouchers, wage subsidies (intervention work), and public works programmes.

The **Spanish** evaluation of the internship contracts also compares the outcomes a well-defined group of programme participants and a pool of eligible non-participants with regression-type models. The treatment group consists of young people under age 30 who have signed an internship contract, while the control group consists of eligible young unemployed people who have signed a regular temporary contract. The sample of social security records enables the construction of employment history variables, such as labour market tenure, number of previous labour episodes, number of previous firms, and time in unemployment. These variables might capture a significant share of the differences between the control and the treatment groups. In addition, the treated and the control groups are defined in a way that allows us to obtain a clean analysis by removing individuals with particular cases from the analysis. For example, one focus is only on those individuals who have just entered the labour market, since the aim of the IC is to place people in entry-level jobs that enhance the stability in the firm. Additionally, those individuals who enter the labour market through an IC are also compared with those who signed an IC after having worked before. This allows us to determine whether the impact of the IC depends on the individual's previous labour market trajectory.

In **Italy**, the evaluation concerns the differences between the eligible and the non-eligible workers in the probability of being hired with an open-ended contract, controlling for the baseline characteristics of the two groups. The main assumption underlying the identification of such an effect via a diff-in-diff methodology is that the incentive does not affect the outcome of the controls; i.e., the policy, which is designed to increase the probability of hiring an eligible worker with an open-ended employment contract, does not affect the employment probability of the non-eligible workers. The descriptive statistics show that this assumption does not hold, as they highlight a discontinuity (reduction) in 2015 in the time series of the outcome variable for the control group. The reason for this discontinuity could be that some employers may have shifted their hiring preferences away from the non-eligible workers (those with higher levels of attachment to the labour market) and towards the eligible workers (those with more fragmentation in their employment careers) due to the new opportunity cost between the two groups caused by the policy¹². In other words, we cannot exclude the possibility that the control group was affected by the treatment itself. This potential bias has been corrected with a two-step procedure. In the first stage, the theoretical outcome values – i.e., the values that would have existed without regulatory changes – are estimated and imputed to the control group. In the second stage, the effects of the hiring subsidies are estimated on the corrected sample using an OLS model with interactions between the treatment and the different age classes in order to retrieve the effects on the young cohorts.

12 As explained in the next section, this issue is made even more complex by the fact that the change in the opportunity cost was uneven among the age groups.

3.4. Heterogeneous effects

An important aspect of the evaluation of a policy is the heterogeneity of the effects of the policy. Does the programme have different effects on different groups – for example, broken down by gender, age, or region – in different years, under different macroeconomic conditions, and with different programme combinations? Understanding the heterogeneities in the treatment effect will help us in understanding and assessing:

- the external validity of the results and the comparability of the results with other evaluation results;
- the cost-benefit efficiency of the programme;
- the weaknesses of the programme and the scope for improvement by adjusting the programme; and
- the gender dimension of the programme.

The programme effects may vary with the level of education of the programme participants. Policymakers will probably welcome the finding that a wage subsidy has a stronger impact on individuals with a lower education, as it may be assumed that this group has less favourable labour market outcomes. Consequently, if the evaluation finds that individuals with secondary or lower education benefit from the subsidy, but that it has no impact on university graduates, the subsidies should be targeted to the former group. Moreover, as it may be assumed that the costs of providing a wage subsidy for a job with higher pay and educational requirements are also higher, targeting lower-educated jobseekers might improve the cost efficiency of the programme. This result implies that subsidies should be targeted to the former group. In Hungary, the 90-day job trial programme, compared with public work found to exert a stronger impact on participants with basic education than on participants with secondary or higher education. This was in line with our expectations, as from the point of view of employment prospects, low educated young people are the most vulnerable group. At the same time, participants of the job trial are better educated on average than participants of public works, indicating that the most employable unemployed young persons are selected into the programme. At the same time, the job trial participants were better educated on average than the public works participants, which indicates that the most employable of the unemployed young people were selected into the former programme. This suggests that higher priority to lower educated young unemployed could increase the average treatment effect while decreasing deadweight loss of the program.

The evaluators may also find that the duration of the programme influences the effects. Such a finding can help to determine how long a similar programme should be in the next period.

The **Spanish** analysis covered the **gender** dimension, and concluded that females were less likely than males to return to unemployment after completing the internship contract. However, like for the other questions under analysis, no statistically differences were found between men and women.

The **Italian** evaluation found significant differences in the effects of the programmes across **age** groups. The results of the difference-in-differences model indicated that the regulatory change had a positive impact on the share of open-ended employment contracts over the total number of employment contracts registered in 2015, decreasing by age class.

The evaluation in Poland found that classroom training and classroom training vouchers worked relatively well for males, but not for females. The public works programmes were shown to be less effective for less-educated individuals and those living in regions with high unemployment. On-the-job training vouchers were found to be more effective than standard on-the-job training, regardless of the gender, the educational level, or the place of residence of the unemployed individuals.

4. ISSUES RELEVANT FOR POLICYMAKERS

Up to this point, we have examined the elements needed to ensure that researchers can estimate the causal effect of the policies (internal validity). However, there are a variety of issues that researchers will want to address when discussing the results of the evaluations with policymakers. Our purpose here is to highlight these questions, and to describe the data needed for potential solutions, without offering detailed guidance on them.



“My question is: Are we making an impact?”

4.1. Interpretation of the results

It is very important for researchers to make clear which effect they are estimating in their evaluation study. In many cases, the evaluation design leads to an internally valid estimation of the average treatment effect on the treated or the local average treatment effect, such as on those who currently take up the measures, or a (small) subgroup thereof. However, when presenting these results to policymakers, it is crucial to emphasise that the results can be generalised to broader populations with additional assumptions. Thus, researchers may want to consider also collecting data on the characteristics of the broader population to help with this generalisation. By contrast, evaluations that aim to estimate the intention to treat (i.e., the intention to offer the measure) or the average treatment effect (i.e., the treatment effect on all those who are eligible, and not just on those who enrol) might produce results that are more directly useful to policymakers, as they incorporate the potential non-take-up (for the ITT), or include the possibility to estimate the causal effect on those who currently do not participate in the measure (for the ATE).

4.2. Link the results to programme design and implementation

Beyond providing a numerical estimation of the programme effects, the evaluation should offer insight into the factors that contributed to the results. For example, if the evaluation concludes that the programme had no significant effect on employment, it is important to know whether the failure was caused by bad programme design (e.g., inappropriate target group, eligibility criteria, financial and personnel resources) or by implementation problems (e.g., lack of information, arbitrary selection of participants).

4.3. External validity, comparability with other studies, and aggregate statistics

Readers of evaluation studies with an interest in policymaking will want to compare the measures that are the subject of these studies with a number of other measures. For example, readers may want to know more about the effectiveness of similar measures for young people in other regions/countries, other (alternative) measures aimed at young people, or the (potential) effectiveness of the evaluated measures for alternative target groups. When presenting such comparisons, researchers should be aware of several issues. First, they should think about how to present their own results so that they can be compared with those of other studies. Second, they should explain how their results can be generalised to other settings in order to discourage simplistic extrapolations, and to draw attention to any limitations. Third, researchers should clearly document the most relevant design elements of the evaluated measures, as well as the prevailing economic conditions, to ensure that other stakeholders are able to understand how these factors might have affected the success of the policies. For instance, the policies evaluated in Hungary, Italy, and Poland were all launched in a recovery period, while the measure studied in Spain could be observed both during a period of economic growth and during the Great Recession.

When interpreting and presenting the results in terms of employment and earnings, policymakers should keep in mind how the given indicator is similar to or different from other well-known or aggregate statistics. For example, the analysis might focus on registered unemployed figures, which can differ sharply from the numbers based on the unemployment concept of the labour force survey. The interpretation should explain these distinctions carefully.

4.4. Displacement effect and deadweight loss

When considering how to provide policymakers with a more complete picture of the effectiveness of measures, two issues should be emphasised. First, a displacement effect takes place if the measure helps young people to find employment, but hurts the prospects of individuals who are very similar to them in skills, but do not have access to the measure. Second, the programme generates deadweight losses if a measure increases the employment of young people, but a non-negligible part of the support provided might not have been necessary, as employers would have recruited these particular individuals even in the absence of the measures. These issues lead to subtle questions being raised about social (political) preferences, or about the relative importance of alternative goals. However, providing evidence on these issues is far from straightforward. Exploring these topics would, at the very least, require researchers to have additional data on other workers at the same and competing firms, and ideally data in which employees are matched to their employing firms.

4.5. Cost-benefit and cost-effectiveness analysis

The approach to impact evaluation we have discussed up to this point has focused on the effects of the policy in terms of labour market outcomes. However, the total programme costs should also be considered, given that policymakers will likely want to compare alternative policies in terms of relative (monetary) yields.

There are two types of analysis that link the costs and the benefits. The cost-effectiveness analysis compares the programme costs with the outcomes expressed in physical units; for example, that the programme created one full-time status for x euros. In the cost-benefit analysis, the effects of the programme are also expressed in monetary units, which allows us to assess whether the total benefits outweigh the total costs. Clearly, this latter analysis is more challenging, since accounting for and expressing benefits requires more information.

Even performing the cost-effectiveness analysis might be quite challenging if the administrative database does not contain the programme costs, which is often the case. The researcher might use aggregate data from other sources to calculate the costs per programme unit. Alternatively, when analysing wage subsidy or apprenticeship programmes, the wage data from a linked social security database, together with the programme rules, could be used to estimate the costs. Thus, researchers should plan to acquire this type of data at the outset of the evaluation.

REFERENCES

Abadie, Alberto and Cattaneo, Matias D. (2018). Econometric Methods for Program Evaluation (August 2018). Annual Review of Economics, Vol. 10, pp. 465-503, 2018
<http://dx.doi.org/10.1146/annurev-economics-080217-053402>

Angrist J.D., Pischke J.S. (2015). Mastering Metrics: The Path from Cause to Effect, Princeton University Press, Princeton.
<https://doi.org/10.1111/j.1475-4932.2011.00742.x>

Angrist, Joshua and Pischke, Jörn-Steffen (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Caliendo M., Mahlstedt R., Mitnik O. (2017). Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labor market policies. Labour Economics, 46, 14-25.
<https://doi.org/10.1016/j.labeco.2017.02.001>

Card D., Krueger A.B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania, American Economic Review, 84, n. 4, pp. 772–793.
<http://davidcard.berkeley.edu/papers/njmin-aer.pdf>

European Commission (2013). Evalsed Sourcebook, Methods and Techniques
https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf

European Commission (2020). How to use administrative data for European Social Funds counterfactual impact evaluations, Directorate-General for Employment Social Affairs and Inclusion, Directorate G.
<https://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8303>

Gertler, Paul J.; Martinez, Sebastian; Premand, Patrick; Rawlings, Laura B.; Vermeersch, Christel M. J. (2016). Impact evaluation in practice (English). Washington, D.C. : World Bank Group.
<http://documents.worldbank.org/curated/en/698441474029568469/Impact-evaluation-in-practice>

Imbens, Guido, and Thomas Lemieux (2008). Regression Discontinuity Designs: A Guide to Practice, Journal of Econometrics, 142, 615-635.
<https://doi.org/10.1016/j.jeconom.2007.05.001>

Leeuw, Frans & Vaessen, Jos (2009). Impact Evaluations and Development: NONIE Guidance on Impact Evaluation. Independent Evaluation Group, World Bank.
<https://www.oecd.org/development/evaluation/dcdndep/47466906.pdf>

Morris, Stephen & Tödting-Schönhofer, Herta & Wiseman, Michael (2013). Design and Commissioning of Counterfactual Impact Evaluations - A Practical Guidance for ESF Managing Authorities. Publisher: European Commission, Directorate-General for Employment, Social Affairs and Inclusion. DOI:10.2767/94454

UNDP (2009). Handbook on Planning Monitoring and Evaluating for Development Result.
<http://web.undp.org/evaluation/handbook/documents/english/pme-handbook.pdf>

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
<https://doi.org/10.1093/biomet/70.1.41>

APPENDIX 1: SHORT DESCRIPTION OF THE EVALUATIONS OF THE FOUR COUNTRIES

ITALY



Programme:

- 1) Lowered firing costs for open-ended contracts (Graded Security Contract)
- 2) Social security rebate for new open-ended contracts and for conversions from a fixed-term to an open-ended position) jointly evaluated.

Data:

The contracts registered in the Ministry of Labour and Social Policies Database (SISCO) referred to the communication sent by employers due by law (COB) between 1 January and 31 December 2015, in contrast to the contracts registered in the previous years.

Outcome:

Share of new hires with an open-end contract over the total of employment contracts registered in 2015.

Identification strategy:

Diff-in-diffs model with a correction for the selection bias (violation of the SUTVA).

Eligible group:

- People hired with an open-ended contract (apprenticeship excluded) during 2014-2015.
- All employers without industrial or geographical specificity. The group includes associations and public enterprises, but excludes public administration and agriculture.
- Employers with no pending contributory arrears or any other irregular situation in terms of collective agreements or territorial agreements.

Non-eligible group:

- People hired during 2015 who had been fired within six months, before the beginning of the new open-ended contract, while employed with an open-ended contract.
- People hired during 2015 who in the three months preceding the reform (01.10.2014-31.12.2014) were not fired by the same employer who had hired them, or by another employer related to the latter, to avoid opportunistic behaviours.

Definition of the outcome variable: the outcome is a dummy, Y , with two possible values: $Y=1$, for open-ended contract; $Y=0$, for temporary contract.

Definition of the treatment variable: $T=1$, for open-ended contract eligible for the treatment; $T=0$, open-ended contract not eligible for the treatment.

The treatment variable is interacted with different age classes in order to retrieve the effects on young cohorts.

Definition of the period: $P=0$, 2014; $P=1$, 2015.

HUNGARY



Programme:

The Hungarian evaluation focuses on the 90-day job trial programme designed for young people under age 25. The job trial provides a 100% wage cost subsidy for 90 days without any obligation to further employment. The programme started in 2015 as a part of the Youth Guarantee Programme.

Data:

The evaluation is based on two administrative datasets. The PES data contain all registered jobseekers and programme participants. The PES data are linked to an administrative dataset of the pension authority, which contains information on the employment status and employment history of about half of the total population.

Outcomes:

The main outcome variable is the employment status 6 and 12 months after completion of the programme, and cumulative wages compared to the minimum wage.

Identification strategy:

The evaluation applies two different identification strategies. The first identification strategy compares outcomes for the 90-day programme participants with those of the participants of other programmes (public works and classroom trainings), applying propensity score matching.

The second estimation is a difference-in-differences specification, and exploits the fact that in the Central Hungarian region, the programme started 9 months later than in other parts of the country due to administrative reasons. The treatment group contains eligible young people in two convergence regions, and the control group consists of eligible young people in the Central Hungarian region. The estimation compares the outcomes of the treatment and the control groups; that is, which individuals registered within 9 months before and after 1 January 2015, when the Youth Guarantee Programme was introduced in the convergence regions. The specification allows for an ITT identification, which shows the effect of the whole YG programme on the eligible subpopulation.

POLAND (EVALUATION 1)



Programme:

A wage subsidy programme for unemployed people under age 30 is evaluated in Poland. It was operating in the 2016-2018 period. The employer-side subsidy of up to minimum wage plus social security contributions was paid for 12 months, and the employers were obliged to prolong the employment for another 12 months after the subsidy expired.

Eligible:

Registered unemployed individuals under age 30 with profile I or II (profile III is non-eligible) were eligible. Employers who reduced employment from this initiative in the last six months were not eligible.

Data:

Data from the public employment services (PES) register with full histories of unemployment registrations and programme participation, not linked to social security data.

Sample:

Individuals who entered unemployment between 2015 and 2016, excluding individuals who were unemployed or participated in an ALMP during the last six months to ensure that we are analysing a new registration that is not a part of a longer unemployment spell. To be able to observe the outcome 36 months after registration, we limit our sample to registrations up to April 2016.

Outcomes:

We analyse two types of outcomes: being off the register 12 to 36 months after registration and the cumulative number of days off the register 12 to 36 months after registration.

Identification strategy:

RDD using the age threshold combined with DiD (2016 versus 2015: other institutional rules change at the age 30 threshold, and they were present in 2015).

POLAND (EVALUATION 2)



Programme:

A set of ALMPs offered to young people: on-the-job training, classroom training, wage subsidies (intervention works), public works, on-the-job training vouchers, classroom training vouchers.

Eligible:

Registered unemployed individuals under age 30 were eligible.

Data:

Data from the public employment services (PES) register with full histories of unemployment registrations and programme participation, not linked to social security data.

Sample:

We restrict our sample to completed ALMPs that started between 2015 and 2016, and lasted for at least one day. We include only participants who were between 18 and 29 years old when the ALMP started, and we exclude individuals who had been in prison. We analyse six types of interventions: on-the-job training, classroom training, wage subsidies (intervention works), public works, on-the-job training vouchers, classroom training vouchers. The total sample includes 319,610 observations.

Outcomes:

Being off the register and not in an ALMP 12 to 36 months after registration.

Identification strategy: Pairwise comparison of programme participants, with propensity score matching.

SPAIN



Programme:

Internship contract. It was implemented in Spain in 1998, and, according to the legislation, the main goal of the programme was to increase labour stability in the area of studies developing professional practices in relation to the level and field of studies, and giving incentives to firms to hire young workers once the period of internship is finished.

Eligible:

The eligibility criteria restrictions refer to education and age: the individual must have finished a bachelor's or vocational training degree and be younger than age 30, or have finished a degree less than 5 years before.

Data:

Continuous Sample of Work Histories (CSWH). This dataset consists of the social security records of 5% of the Spanish population, and the complete labour market histories of a representative sample of the Spanish workers.

Sample:

Individuals who were under age 30 and highly educated (i.e., potential users of the IC) who entered employment from 2002 onwards.

Outcomes:

We analyse two types of outcomes: i) the probability of remaining at the firm after the IC and ii) the probability of finding a permanent contract after the IC, in particular at the same firm (for those who remain) or at another firm (for those who change).

Identification strategy:

For the first analysis, we develop a multinomial logit regression in which the dependent variable takes three different values: specifically, i) change to another firm, ii) remain at the firm, and iii) enter unemployment. In the second step, we estimate separately the probability of signing a permanent contract (compared to signing a temporary contract) depending on whether the individual remains at the firm or changes to another firm with a logit estimation model.

APPENDIX 2: GLOSSARY AND ABBREVIATIONS

ALMP, Active labour market policy

ATE, ATT, average treatment effect and average treatment effect on treated: A treatment effect is the causal effect of the treatment (a binary, 0–1 variable) on an outcome variable. It captures the difference between the potential outcome of a population unit with and without the treatment (exposure to the policy, taking part in a specific programme, etc.). There are two major concepts of the average treatment effect. The ATE shows the population expectation of the treatment average difference in the pair of potential outcomes averaged over the entire population of interest. This is the relevant measure if the entire population is exposed to the policy under consideration.

$$ATE = E(Y_i(1) - Y_i(0))$$

where $Y_i(1)$ is the outcome of the unit i when she receives the treatment, and $Y_i(0)$ is the outcome of the unit i when she does not receive the treatment.

The ATT, or the average treatment effect on treated, shows the average of the treatment effect over the subpopulation of the treated:

$$ATE = E(Y_i(1) - Y_i(0) | D_i = 1)$$

CIE, counterfactual impact evaluation

CO, COB, Comunicazioni Obbligatorie online, registry of mandatory communications managed by the Ministry of Labour and Social Policies (Law n. 296, 27 December 2006 (Financial Law 2007)).

CSWH, Continuous Sample or Work History, a 5% sample of Spanish social security records.

Diff in Diffs, DiD, difference in differences: A quasi-experimental estimation method in which the effect of the programme is calculated by comparing the treatment group and a control group before and after the treatment. The treatment effect is the average change over time in the outcome variable for the treatment group, minus the average change for the control group. The DiD method relies on the assumption that in the absence of the treatment, the outcome of the treatment group would have followed the trend of the untreated units; that is, the two groups would have had parallel trends (for more details, see, e.g., Card and Krueger, 1994, Angrist and Pischke, 2008).

ISCED, International Standard Classification of Education.

https://ec.europa.eu/education/international-standard-classification-of-education-isced_it

INAPP, Istituto Nazionale per l'Analisi delle Politiche Pubbliche

ITT, intention to treat: This indicator shows the effect of the policy on the eligible population. In other words, it shows us the causal effect of the offer of treatment. If not all members of the eligible population receive it, as many of them will decline it, the ITT will differ from the average treatment effect.

NACE, Statistical Classification of Economic Activities in the European Community, Rev. 2 (2008), Rev. 2 (NACE Rev. 2) https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2&StrLanguageCode=IT&IntPcKey=&StrLayoutCode=HIERARCHIC

OLS, ordinary least squares

PES, public employment service

Propensity score matching: This is a counterfactual evaluation method that compares the outcome variable for treated individuals with the outcome variable for matched individuals in a control group who are similar in their observable characteristics to the treatment units. In propensity score matching, the treatment units are matched to control units with a similar propensity score; that is, the probability of being treated given a set of observable variables. This method relies on the assumption that conditional on observable characteristics, selection into the treatment group is random (for more details, see, e.g., Rosenbaum and Rubin, 1983; Angrist and Pischke, 2008).

RDD, regression discontinuity design: This is a quasi-experimental method that allows us to estimate the treatment effect of the programme when the eligibility for or the probability of participating in the programme depends on a certain observable characteristic, such as age (forcing, or running variable). The effect of the programme is measured by comparing the outcomes below and above the threshold in the close neighbourhood of the threshold. The RDD calculates the local average treatment effect by fitting a (local) polynomial on both sides of the threshold, and relies on the assumption that the selection between the treatment and the control groups is random. RD can be fuzzy and sharp (for more details, see, e.g., Imbens and Lemieux, 2008; Angrist and Pischke, 2008).

SISCO, Sistema Statistico delle Comunicazioni Obbligatorie online, mandatory communication collected by the Ministry of Labour and Social Policies (MLSP).

SUTVA, stable unit treatment value assumption: This is an assumption that is usually made in causal inference: i.e. a treatment applied to one unit does not affect the outcome for another unit (no interference between the units).

Unconfoundedness: This term refers to a case in which controlling for differences in a fixed set of pre-treatment of covariates removes biases in comparisons between treated and control units, such that conditional on those controls, the treatment assignment can be regarded as randomised. Also known as ignorability or the conditional independence assumption.

© BUDAPEST INSTITUTE FOR POLICY ANALYSIS 2020

Address: Dohány utca 84, Budapest, 1074 Hungary

info@budapestinstitute.eu

www.budapestinstitute.eu



This work is licensed under a [CC BY-NC 4.0 license](https://creativecommons.org/licenses/by-nc/4.0/).

ISBN 978-615-01-0433-1



9 786150 104331