Impact evaluation of labour market and social policies through the use of linked administrative and survey data

Technical Report: Impact Evaluation of Ireland’s Active Labour Market Policies

March 2024

An Roinn Coimirce Sóisialaí
Department of Social Protection

Please cite as:
OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre (2024), “Technical report: Impact Evaluation of Ireland’s Active Labour Market Policies”, OECD, Paris, www.oecd.org/els/emp/Ireland_ALMP_Technical_Report.pdf.

Co-funded by
the European Union

This work was produced as part of the project “Pilot studies on impact evaluation of labour market and social policies through the use of linked administrative and survey data” which is co-funded by the European Union (European’s Commission Directorate-General for Employment, Social Affairs and Inclusion) (VS-2020-0 368).

The views expressed herein can in no way be taken to reflect the official opinion of the European Union.

Specific territorial disclaimers applicable to the OECD:

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Table of contents

FIGURES

1 Executive Summary

This report is produced in the framework of a project of the OECD with the European Commission (EC) which aims to raise the quality of the data collected and their use in the evaluation of the outcomes and effectiveness of labour market programmes, so that countries can better evaluate and design policies to benefit their citizens.[1] Within the OECD-EC project, a joint project team was established with representation from the OECD, the Department of Social Protection (DSP) and the Joint Research Centre of the European Commission (JRC). The joint analytical team conducted a counterfactual impact evaluation (CIE) of Community Employment (CE), an impact evaluation of Tús and an analysis of the sequencing of ALMPs. The results and the consequent policy recommendations are published in the OECD series *Connecting People with Jobs* (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]). This technical report accompanies the main report on evaluation results, aiming to build analytical capacity and inform future CIEs conducted by the Irish authorities. The technical report includes a detailed discussion on the administrative data available for the evaluation used in the main report. It draws insights on the process undertaken to extract and compile these data, so that future impact evaluations may build further on the work necessary for this report. Further details are provided on the analysis undertaken, to supplement the main report and to provide additional scrutiny and assurance of the techniques used in that report. The report also includes supplementary analysis, offering additional insights and elaborating on the methods and results of the evaluations conducted.

The data available for use in these reports have been compiled by analytical teams in the DSP, drawing data across a number of administrative systems. Broader register data on receipt of unemployment benefits and utilisation of Revenue Ireland data on earnings, enable the reports to build a rich picture of the individuals participating in active labour market policies (ALMPs) and on their experience in the labour market. In addition to these core data, data on DSP's risk scoring questionnaire, the Probability of Exit (PEX), is available for a sub-set of jobseekers, with coverage greater for more recent cohorts. These data provide an interesting additional set of information that permits the reports to look further into different sub-groups and to assess the extent to which omitted information on individuals may have the potential to bias the conclusions in the central analysis. As the data are limited to those held by DSP, there are limitations on the extent to which the analysis can provide information on broader outcomes, for example on health or take-up of education places. This is of particular concern for the analysis of CE, which has an explicit objective on "social inclusion". The reports utilise information on wider DSP benefit receipt to proxy these as best they can.

The methodology for the impact evaluation of CE in the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) was determined by the availability of rich administrative data and by the selection criteria for the ALMPs. For CE, where jobseekers have discretion over whether or not to participate in the programme, matching techniques are utilised.

[1] "Pilot studies on impact evaluation of labour market and social policies through the use of linked administrative and survey data" which is co-funded by the European Union (European Commission's Directorate General for Employment, Social Affairs and Inclusion) (VS 2020 0368).

These seek to ensure that characteristics that influence both the choice of jobseekers to enter the programme and their subsequent labour market experiences, are similar between the participants and the non-participants to which they are compared to. The idea is to ensure that the only potential difference between the comparison groups is related to the participation in CE, rather than underlying innate differences of individuals.

For Tús, in addition to self-selection of participation similar in CE (for 30% of participants), there is a randomised referral system in place (for 70%). In principle this means that Tús functions akin to randomised control trial for these individuals. As selection for referral takes place randomly, it should guarantee that differences between participants and non-participants are only attributable to Tús participation.

However, the extent to which the reality of random selection for Tús and compulsory participation occur in real-life, limits the ability of the analysis to proceed as if it were a randomised-control trial. There are aspects of both administrative selection and self-selection inherent in the journey from referral to commencement. Therefore, similar techniques to those used in the analysis of CE need to be employed to ensure that unbiased estimates of programme impacts are produced.

The analysis of sequences of participation in ALMPs proceeds in a slightly different fashion. It arranges data so that the main pathways for jobseekers are in terms of the ALMPs that they participate in. This analysis can provide important extra information for the separate CE and Tús evaluation, demonstrating how individuals move from one to the other.

A natural next step for the sequence analysis would be to ascertain whether causal pathways are identifiable. For instance, participation in CE might yield better or worse outcomes if it is preceded by participation in Tús. The standalone analysis of CE would not be able to determine this distinction and would instead produce an aggregation of outcomes from individuals that had and had not previously participated in Tús. This causal analysis is beyond the scope of this technical report.

All of the choices on methodology for the evaluations of CE, Tús and sequences of ALMPs contained within this report confer assumptions, which if they are violated mean the impacts do not truly describe the impacts that they estimate. This technical report therefore describes additional analytical checks that have been performed in order to provide additional scrutiny on the robustness of these results and their sensitivity to changes in the analytical specification and techniques used.

This report now proceeds to outline the data available, the processes and choices made for linking and compiling the datasets for evaluation and the additional analytical checks that have been made to assess the results stemming from the central analysis presented in the main report.

The report leads to some findings that could also support future evidence building and operational delivery for ALMPs in Ireland, with improvements that could be made to its data, including data gathering, storage and availability:

- Build on the extensive work undertaken for this project, to compile and maintain a consistent analytical dataset that can be used for future impact evaluations. This will ensure both efficient and timely analysis and will provide a consistent framework across evaluations.
- Review the process of recording job descriptions in both CE and Tús to ensure alignment. This will serve dual purposes. From an analytical perspective it will ensure easy mapping between the two similar schemes, so that comparative analysis can be more readily performed. For individuals participating in the schemes (for whom there is a large cross-over between the two), it will mean provide greater clarity and consistency on the roles available and the expected tasks across the programmes.
- Ensure that PEX data are captured for all live register claimants for that these rich data can inform policy delivery in Intreo centres and inform future policy evaluations.

- Capture administrative data on ‘social inclusion’ objectives for CE, either directly in administrative systems or via the use of client surveys. This will permit scrutiny of whether or not CE is successful in achieving its aims in this regard. In addition, gather data on education.

2 Administrative Data Sources

2.1. Data linking and preparation

DSP is one of the largest holders of personal data in Ireland. In order to protect these data, the Department's Information Services Division has put in place an Information Security Management System (ISMS) with a set of controls which include policies, processes, procedures, organisational structure, and software and hardware. DSP currently operates IT security management in line with ISO27001 international standards.

Its controls are monitored, reviewed and updated on a regular basis to ensure that they protect information assets regardless of how the information is formatted or whether it is in transit, is being processed or is at rest in storage.

The data used to conduct the impact evaluations in the Irish project come from several sources and some of these data originate outside of DSP. This section sets out how these data were linked and what cleaning and preparation was required in advance of any analysis.

The first point to note is that there is no single database of administrative data that has been designed for analytical purposes. The Jobseekers Longitudinal Dataset (JLD), which collated unemployment, training and employment episodes has not been maintained since 2018. While useful for some of the initial work establishing the eligible population, the JLD could not be used for an evaluation that seeks to examine outcomes up to 2021. Censored dates for claims in the JLD had to be compared to more recent register data to update end dates and establish claim duration. In the absence of an analytical dataset that could be used for the analysis of CE and Tús and the sequences of long-term unemployed people, DSP collated a series of individual-level datasets built from operational tables.

Individual level datasets in DSP can be linked using a unique individual identifier. In Ireland, the Personal Public Service Number (PPSN) is a unique reference number to enable access to social welfare benefits, public services and information. The presence of this variable on all of the individual level DSP and Revenue Commissioner datasets means it is a straightforward step to join different information about an individual from several sources.

However, while this illustrates the benefits of a single public service identifier, it also bears repeating that great care was taken when accessing, linking and extracting data. In preparing the administrative data, steps were taken to remove extraneous data, aggregate the data by using month and year of birth in place of date of birth, removing identifiers and pseudonymising the data. Unique identifiers that have application beyond the needs of linking datasets for analytical purposes, were removed. Instead of a PPSN, a pseudonymised ID was generated to link data. The consistent use of this identifier enables individuals to be easily linked, in a secure manner, and in a way that precludes this linking variable to identify the individual in any way. The PPSN was not included in any datasets used in this analysis.

The DSP also carried out a Data Protection Impact Assessment to ensure the pseudonymisation, data minimisation and simplification steps were sufficient to minimise and mitigate any risks identified. This was done in accordance with the seven privacy principles which have been set out by Article 5 of the GDPR, which incorporate all the data protection and privacy requirements within EU and Irish legislation. In addition, the analysis proceeded in accordance with the CSO Best Practice for Statistical Disclosure Control of Tabular Data.

2.1.1. The construction of the participant and control group populations can explain some variation in published statistics in the analysis.

As is noted in section 2.2 below, members of certain sensitive groups who take part in CE/ Tús are excluded from the analysis. In addition to this exclusion, decisions around what qualifying claims and eligibility routes should be considered in the eligible population were made when constructing the datasets. CE for example has a wide range of qualifying claims but not all are well represented in the participant population. As a result, the number of participants in both CE/ Tús reported in the analysis may show some variation from published statistics.

The process of identifying the eligible population begins with the existence of a qualifying claim. These are slightly different for Tús and CE, with Tús having a narrower focus on people who are fully unemployed (i.e., not in part-time employment and not only in receipt of credited social insurance contributions) for at least 12 months and CE extending to people in receipt of One Parent Family Payment (OFP). For the participants, the qualifying claim was defined as the claim which the individual closed within 90 days of starting CE/ Tús.

As the qualifying age for CE for those on the live register reduced from 25 to 21 years in 2017, eligibility is calculated to identify the later of turning 21 or becoming long-term unemployed after 2017. Similarly, lifetime participation limits changed over time and are dependent on the stream participants are on and their age. Eligibility is calculated each quarter according to age and the rules in place at the time with regard to age limits and lifetime participation limits.

During the analysis period, the option for recipients of Jobseeker's Transitional payment (JST) to self-refer to Tús was introduced. JST is a payment for people parenting alone whose children are aged between seven and 14 years. As many have been recipients of OPFP previously, the JST cohort bears a closer resemblance to OPFP recipients than to people in receipt of the benefits counted as part of the Live Register. Their inclusion in the main analysis would necessitate the inclusion of JST recipients in the control group despite the limited number of whom actually self-refer to Tús. In the period of analysis, participants from the self-referral route made up approximately ~4% of all participants. As a result, all participants who self-referred (either from a jobseeker or JST claim) were excluded from the analysis and the focus is solely on the difference in labour market outcomes between long-term recipients of Jobseeker's Allowance or Jobseeker's Benefit who participated in Tús and those who did not.

Box 2.1. Data storage within DSP

Data from the Department's operational databases are stored on various servers and arranged into schemas within databases on each server. Due to the volume of data held by the Department and the absence of one single analytical database, assembling the datasets for this evaluation was the complex task. It begins by linking each relevant SQL table to construct an episodic history for each individual. While the existence of a single public service identifier facilitates data linking across different sources (such as data collected by the Revenue Commissioners and each DSP database), the joining of data within each schema presents its own challenges.

A good example of this challenge arose when constructing the datasets with information on CE participants scheme activities, job titles and education and training records. All of the data come from the same database schema (and server) which allowed for merging across tables. However, not many of the tables needed for the analysis contained the internal DSP individual identifier, which necessitated joining on different variables, which were only present in some tables. Often, intermediate tables, that did not include variables for analysis, were included in the merge just to facilitate linking of required variables. An example of this is illustrated in Table 2.1 in assembling data on training records for CE participants.

Table 2.1. SQL schema structure

individual_identifier	outcome_id	activity_id
activity_id	end_date	training_award_level_code
Outcome_id	Outcome_reason_code	Training_award_type_code
Individual_identifier	outcome_id	activity_id
Activity_id	end_date	Training_award_level_code

In the above table, the variables needed for analysis are contained in Table B and C (end_date, outcome_reason_code, training_award_level_code and training_award_type_code). However, Table A is the only table that includes the individual_identifier which can link training records back to each person. Table A is needed in the data linking phase as an intermediate table as it can (i) link the data to an individual and (ii) link Table B and Table C who otherwise do not share a common variable. Another feature of the schema structure is highlighted in Table X.1 through the 'code' variables. Look-up tables like Table 2.2 containing just the code variables and the corresponding descriptions are common within schemas.

Table 2.2. Schema look-up table

Training type code	Description
HSA	Health and Safety
JRT	Job Related Training

While data minimisation was a priority when assembling the data for transfer, descriptive tables such as Table 2.2 were crucial in explaining variables, especially in the absence of complete metadata and for a project such as this where researchers external to DSP were involved.

2.2. Data sources

The final data used for the analysis in the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) are generated by linking of up to ten different administrative datasets. The core data set used for the impact evaluation of Community Employment (CE) is mostly determined by the dataset containing individuals who were eligible for CE (described in 2.2.1), the Tús evaluation similarly determined the eligible Tús population, whilst the sequencing analysis makes use of the full suite of individual-level datasets available. The following section gives an overview of the input data sets and presents some descriptive statistics.

2.2.1. CE eligibility dataset

The CE eligibility dataset contains individuals who were eligible for CE at some point between 2013 and 2018. The data are organised according to spells of eligibility. Therefore, individuals can occur several times in the database if case they were eligible for CE on more than one occasion.

The vast majority (92.5%) of eligibility spells in the data start at between 2009 and 2018, but in some cases, people have remained eligible for CE for a long time. These are people who have remained long-term unemployed for several years and did not participate in another active labour market programme preventing them from being eligible for CE. About 1% of eligibility spells start in 2004 or earlier.

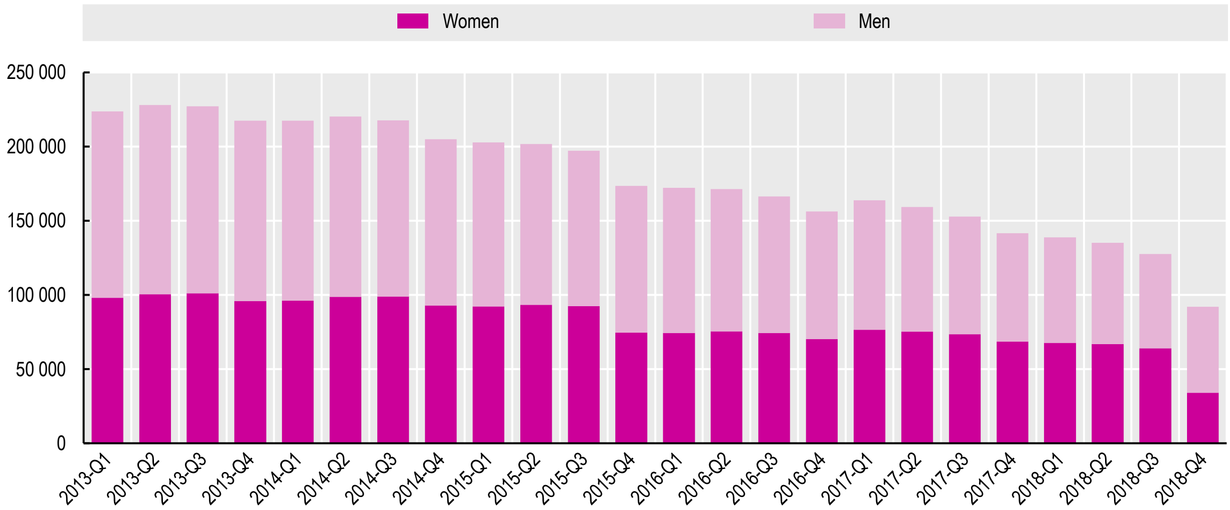
The length of eligibility spells, which typically start 12 months after a person became unemployed, and end as soon as a person is no longer available for CE, is very diverse. While 10% of spells last less than two months (60 days), the longest 10% last almost 6 years or longer (2116 days or more). The median length of an eligibility spell has a duration of 473 days.

In total, there are 721 202 eligibility spells in the data, corresponding to 630 255 distinct individuals, i.e., most individuals were eligible for CE only once during the observation period. For people with more than one eligibility spell, the average time between two eligibility periods was just above two years (752 days).

The number of people eligible for CE decreased between 2013 and 2018, reflecting marked labour market improvements over this time period as the repercussions of the European debt crisis started to easy (Figure 2.1). At the beginning of 2013, more than 220 000 people were eligible for CE, against just above 90 000 at the end of 2018.

Figure 2.1. The number of people eligible for CE dropped sharply after the end of the European debt crisis

Number of eligibility spells in a given quarter, by sex



Note: In some few cases, one individual can have more than one eligibility spell in the same period.
Source: Department of Social Protection (DSP) data.

2.2.2. CE participant dataset

The CE participant dataset contains information on jobseekers who started CE between 2007 and 2020, and in addition a very limited number of participants who started in either 2006 or 2021. The dataset is organised according to CE spells, that is, each individual can be included several times in case they held at least two different CE positions over this period. After correction for mistakes in the data set (e.g., CE spells with negative length etc.), there are 340 961 CE spells in the data which correspond to 96 892 individuals. Stated differently, each CE participant has an average of 3.5 CE spells.

Table 2.3. Many Community Employment participants have more than one participation spell

Distribution of the number of Community Employment participation spells per individual

Number of Community Employment participation spells	Frequency
1	13%
2	24%
3	20%
4	21%
>4	22%

Source: Department of Social Protection (DSP) data.

Almost half (46%) of CE spells last almost exactly one year (between 350 and 375 days), while 15% are shorter than 100 days and 36% fall in the range 100-350 days. CE spells with a length of more than one year also exist in the data but are uncommon.

In terms of job types, 95% of CE spells correspond to standard CE participant while 1% are assistant supervisor roles and 4% supervisor roles. This pattern is in line with the standard approach to employ one (assistant) supervisor for a group of about 20 people. Almost all CE spells (96%) are classified as mainstream CE, with the remainder being categorized as dedicated childcare schemes (2%), dedicated health and social care (2%) or other sparsely populated schemes (e.g., drug rehabilitation with 0.06% of the data set).

The data also include some information on the participant's whereabouts about the end of the scheme, although not always the correct outcome may be recorded in the data. Upon completion of CE, more than two-thirds (69.9%) of participants are transferred to another scheme, which could be another CE assignment or participation in some other ALMP. For the remainder, 13.7% are recorded as "unemployed" while 8.1% are recorded as having found employment after CE participation.

For some CE spells, the data also includes a description of the specific role within the CE scheme. Among the most common job descriptions are childcare workers, caretakers, environmental workers and more generic descriptions such as general operatives and administrators. The data extraction of job descriptions required text cleaning to match job descriptions to the categorization of 'Social Inclusion' or 'Activation' by searching for key words from a job title, correcting for whitespace and case differences, or removing additional text.

For some of the CE participants and CE spells, additional data are available on educational attainment and training during CE. More precisely, out of the 96 892 individuals who took part in CE in the data, for 71 624 there is information on their highest educational attainment. Almost half of them (47.1%) have education no higher than National Framework of Qualifications level 3 (up to Junior Certificate), and less than one-fifth (19.2%) have education above level 5 (higher certificate, advanced certificate or above). Less than 2% hold a master's or doctoral degree. The data do not contain information on when the degree was earned, i.e. the education level may have been obtained before or after CE participation.

Regarding training during CE, among the 340 961 CE spells in the data, 50% were supplemented by at least 2 days of training and 25% by at least 58 days of training. These numbers disregard briefing sessions as well as training that was cancelled, finished early or unsuccessful. Among the CE spells for which detailed information on training is available, the most common training involved minor components/awards (46.2%) and not-certified training (35.2%), while only a minority of CE spells were accompanied by training for a Quality and Qualifications Ireland (QQI) major (5.3%). The median number of days in training was much higher for CE spells with training leading to a QQI major (171 days) than for a QQI minors (58 days) and non-certified training (16 days). In addition to training for job-specific skills, ranging from care for older persons to forklift training, training on "Health and Safety" was common.

Table 2.4. There are diverse training programmes for CE participants

Common training titles, by training award types

	QQI Major	QQI Minor	Non certified
Most common training title	Occupational first aid	Occupational First Aid	Manual handling
2nd most common	Childcare	Pesticide Application	Safe pass (safety awareness training programme for construction workers)
3rd most common	Healthcare support	Horticulture Tools and Equipment	Health & Safety

Note: The list only accounts for courses with the exact same title. Therefore, the list should be interpreted with caution.
Source: Department of Social Protection (DSP) data.

The number of training days tended to decrease somewhat over time. Among those starting CE in 2009 or 2010, the 75th percentile of training days was about 105 (i.e., 25% of CE spells were accompanied by at least 105 training days), against about 53 days for those starting in 2014 or 2015 and 31 days for those starting in 2019.

There is no systematic and apparent link between training participation and other important variables in this data set, such as outcomes after CE participation (as recorded in this data set) and job descriptions.

2.2.3. Tús eligibility dataset

The Tús eligibility dataset contains individuals who were eligible for Tús at some point between 2013 and 2018. Like CE eligibility, the data are organised according to eligibility spells.

The vast majority (94.6%) of eligibility spells in the data start at between 2009 and 2018, but in some cases (as in CE eligibility), people have remained eligible for a long time. About 1.5% of eligibility spells start in 2004 or earlier.

The length of eligibility spells, which typically start 12 months after a person became unemployed, and end as soon as a person is no longer available for Tús, is very diverse. While 10% of spells last less than 1 months (27 days), the longest 10% last almost 5 years or longer (1867 days or more). The median length of an eligibility spell has a duration of 354 days.

In total, there are 534 934 eligibility spells in the data, corresponding to 463 463 distinct individuals. It is important to note (notably for the sequence analysis) that there is a large cross-over between Tús eligible individuals and CE eligible individuals (in most cases individuals will be eligible for both simultaneously, therefore the distinction in the analytical datasets is largely one of analytical convenience). For example, of the 463 463 Tús eligible individuals, 396 166 also appear in the CE eligible dataset. Most people were eligible only once for Tús (86.2%), whilst 12.3% had two eligibility spells and 1.4% three.

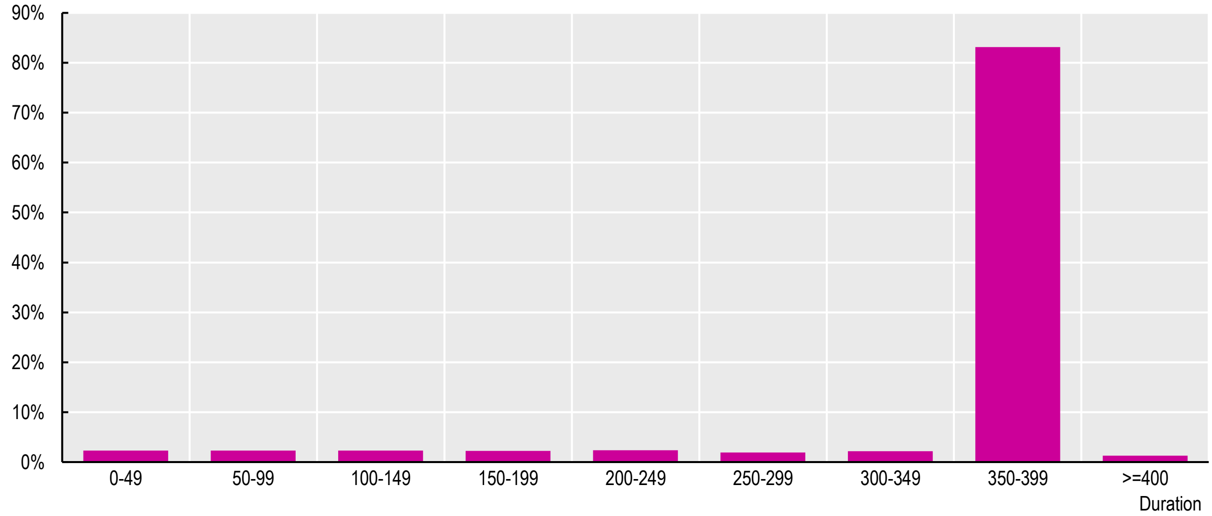
Similar to the CE eligibility data, the Tús eligibility dataset contains individual level information on dates of birth and death (suitably pseudonymized), the benefit payment rate associated with the individual, child dependents allowance, previous occupation, total prior contributions until date of eligibility and variables that detail previous live register status types.

2.2.4. Tús participant dataset

The dataset on Tús contains 44 045 participants. Most of them have a record for only one Tús participation, but 1 206 individuals participated twice in Tús and 12 even three times. The low rates of multiple participants are partially explained by the restriction that former participants can only enrol in a second Tús programme after a waiting period of three years (Government of Ireland, 2023[2]). Four out of five (83%) of Tús participants leave Tús because the programme ends, while 6.5% leave because they find full-time employment and the remainder for less common reasons, such as health problems (1.9%), enrolment in education/training (1.3%) or because Tús participation is considered as not financially viable (0.4%). The most common work types are classified as Community development (21%), working in sport facilities (14.8%) and working in charity shops (9%). In line with the exit reason, 83% of participants record a duration of almost exactly one year (350-365 days), while 8.5% leave the programme after less than half a year (182 days). For the 1 206 individuals with more than one Tús spell, the median length between two spells was three years (1116 days), but 16% of them started the second Tús position directly after ending the first one, which is possible if the first spell ended prematurely.

Figure 2.2. The vast majority of Tús participants leave the programme after 1 year upon completion

Frequency of the length of Tús participation, in days



Note: The graph reflects a person's first Tús participation and does not take account of the limited number of cases in which one individual participates more than once.
Source: Department of Social Protection (DSP) data.

2.2.5. Earnings dataset

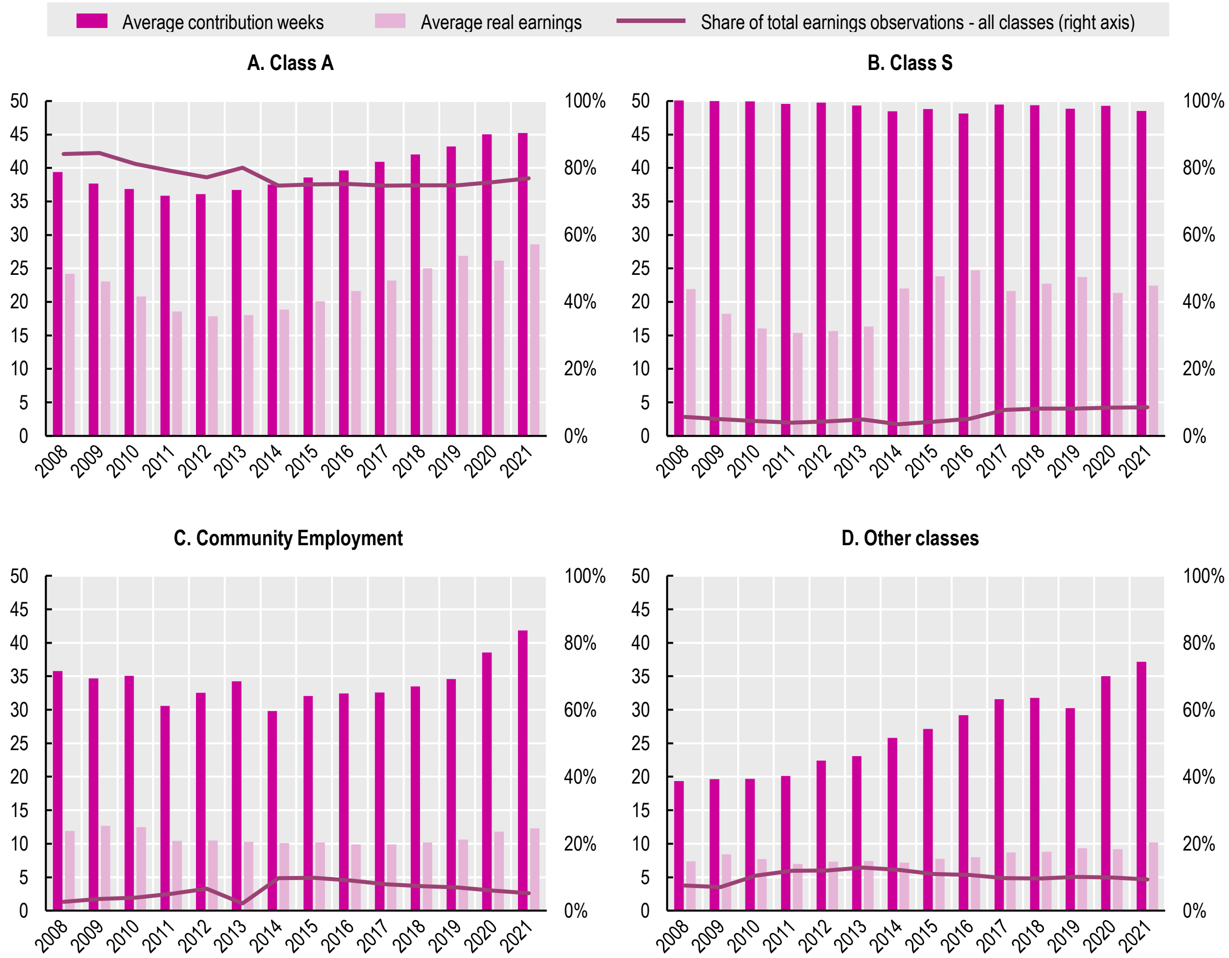
The earnings data set contains information on earnings from different social insurance classes between 2008 and 2021, enabling the tracking of income of potential CE and Tús participants before, during and after their eligibility. More specifically, the data distinguish between earnings stemming from:

- Social insurance class A: People employed in the private sector earnings at least EUR 38 a week and civil servants recruited from 6 April 1995 (Government of Ireland, 2023[3])
- Community Employment (social insurance sub-classes A8/A9)
- Social insurance class S: Self-employed people including farmers and sole traders, investors.
- Other types of earnings (generic term encompassing some other types of pay-related social insurance classes)

Average real earnings of CE participants are, on average, slightly less than half of that of employees with class A income, and slightly more than half of average class S earnings (Figure 2.3). While real earnings of employed and self-employed people dwindled during the European Debt crisis, falling sharply until they hit a trough between 2011-2013, and the recovered, real earnings from CE were far less volatile. The average number of contribution weeks towards CE is somewhat lower than for class A, and even more so than for class S, but it has been increasing over the last few years, from 32 weeks in 2016 to 42 weeks in 2021.

Figure 2.3. Real earnings of CE participants amount to about half of earnings of employed workers and are far less volatile

Average real earnings (in EUR 1000), average contribution weeks (in weeks) and share of earnings observations of specific class among all earnings observations (in %), by income class and year

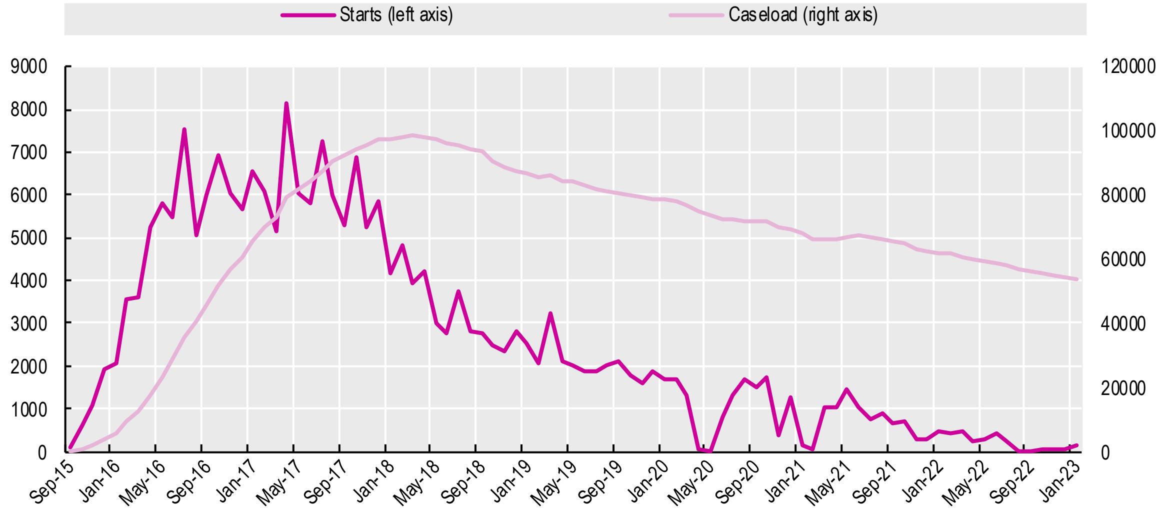


Source: Department of Social Protection (DSP) data.

2.2.6. Job Path data

The Job Path dataset contains information on referrals to the contracted-out employment service Job Path, with start and end dates for individuals. The dataset has 230 487 observations, that cover 156 012 unique individuals. Job Path starts occur between September 2015 and January 2023. Most individuals (60%) have only one referral, 33% have been referred twice, 6% three times and only 0.5% have been referred more than three times. Starts to Job Path were highest in 2016 and 2016, averaging 5 700 per month, before a continual reduction across the rest of the time horizon to fewer than 100 per month by August 2022 (Figure 2.4). The inclusion of Job Path data allows the CE and Tús analysis to account for restrictions to eligibility that Job Path entailed- so that individuals undergoing Job Path during some of the analysis period were not eligible to start either CE or Tús, until the completion of their Job Path programme.

Figure 2.4. Job Path starts and caseload by month

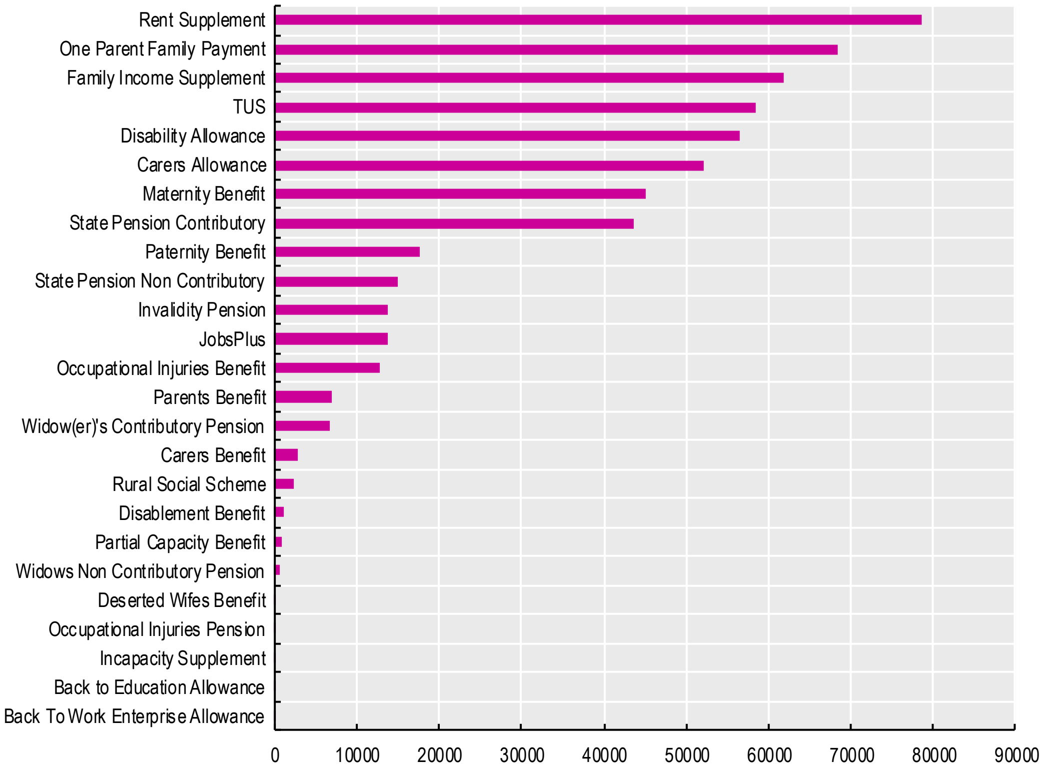


Source: Department of Social Protection (DSP) data.

2.2.7. DSP scheme dataset

The DSP scheme dataset provides information on receipt of different DSP administered benefits. What are the years 2008-21. It is a spells dataset, so that each observation contains information on start and end dates of discrete periods of receipt of a benefit (for example, an individual with two separate periods of receipt of disability benefit would have an observation recording the start and end dates for each of those periods). The most populous benefit for individuals in that dataset is Rent Supplement with 78 600 separate occurrences (Figure 2.5), followed by two family benefits (One Parent Family Payment and Family Income Supplement

Figure 2.5. DSP scheme data

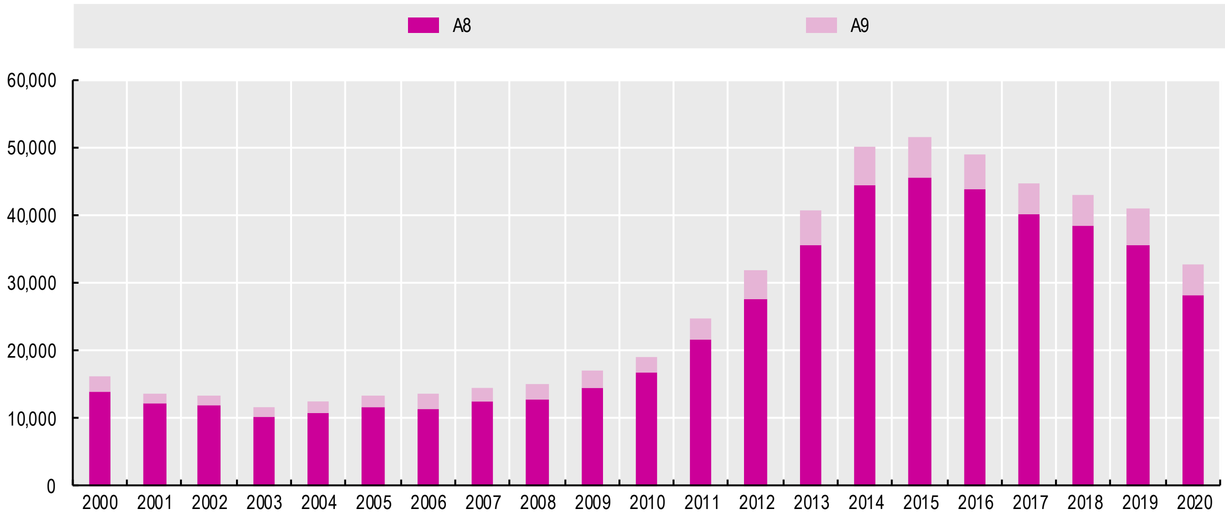


Note: Pandemic Unemployment Payment excluded from chart (184 433 observations). Sorted from largest to smallest.
Source: Department of Social Protection (DSP) data.

2.2.8. A8 / A9 contribution dataset

The A8/A9 contribution dataset contains annual records on the payment of CE social contributions. The distinction between A8 or A9 contribution depends on the amount of weekly earnings. In 2023 the bands were CE pay of up to EUR 353 per week for A8 contributions and over this amount for A9 contributions (DSP, 2023[4]). The dataset has records going back to 2000, of which 88% are A8 contributions (Figure 2.6). The sample frame for this dataset is all individuals identified in the CE eligible dataset. For this reason, we see a peak in observations around 2015 (at 52 000 total observations), as our sample for eligible individuals are those individuals eligible for CE in the period between 2013 and 2018. These observations are used to compute lifetime limits on CE participation for the CE analysis.

Figure 2.6. A8 / A9 contributions by year



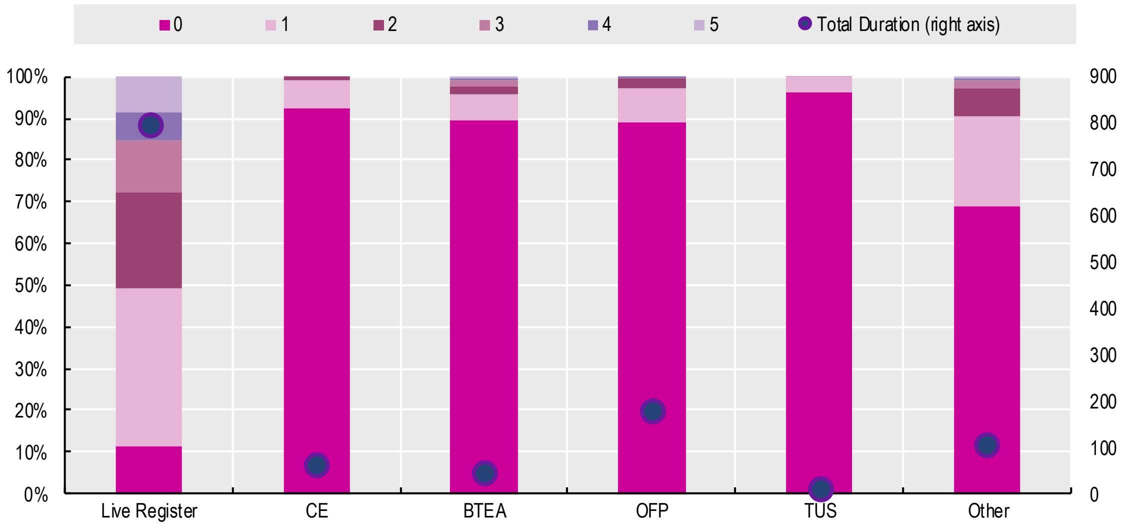
Source: Department of Social Protection (DSP) data.

2.2.9. Pre-qualifying episode dataset

The pre-qualifying episode data summarises individuals previous benefit histories. It is linked to the CE eligibility dataset in that an individual has a record created for every qualifying episode they have. In this way the dataset provides an up-to-date collation of an individuals' past benefits at the point at which they became eligible for either CE or Tús. It has 816 456 observations, which cover 432 899 unique individuals. Thirty eight percent of individuals have one record, 47% have two records, 7% have three, 7% have four and 2% have over four.

The most populous previous benefit by far is the Live Register (Figure 2.7). The mean number of Live Register episodes per individual is 2, but 15% have had 3 or more spells on the Live Register. This is reflected in the average duration as well, with the Live Register mean being 796 days. The next biggest benefit, in terms of the mean length of receipt is One Parent Family Payment (OFP) at 178, but this benefit is no-where near as widespread. Eighty Nine percent of individuals have 0 spells of OFP. As Figure 2.7 shows, Community Employment (CE), Tús and Back to Education Allowance (BTEA) have a similar distribution of individuals without any receipt (92%, 96% and 90% respectively).

Figure 2.7. Distribution of benefit episodes by frequency



Source: Department of Social Protection (DSP) data.

2.2.10. Demographic dataset

The demographic dataset contains information on individual's characteristics. It has 1 119 326 records of 704 163 individuals. Just over half of individuals (53.1%) have only one record, 36.6% have two and 9.0% have three, a further 1.5% have four or more. The dataset contains information on gender, marital status, broad nationality and location. Individuals with more than record are usually those with an updated marital status (changes to marital status account for 236 205 of observations). A further 178 899 records account reflect observations that are updated only in the dates in which they were created, so are not substantive in the sense of changes to either marital status or location.

2.2.11. Excluded categories

Finally, CE and Tús participants potentially include members of sensitive groups, such as people with a history of drug misuse, people who have been convicted of crimes or members of groups that face discrimination such as members of the Traveller and Roma Communities. People who are participants on CE schemes where participation indicates a particular racial or ethnic origin, or where it reveals information about a person's health, were not included in the evaluation. No data associated with participation in schemes aimed at these particular cohorts were analysed and the individuals were not included as potential comparison group candidates if they later became eligible (or had previously been eligible) through other routes (e.g., as a jobseeker). This aimed to eliminate the risks associated with sensitive data, such as data relating to racial or ethnic origin, or data concerning health.

2.3. Data Limitations

Evaluating the impact of ALMPs requires rich data with detailed information on jobseekers' characteristics, their labour market histories, their participation in a range of ALMPs, their previous receipt of income support and their engagement with employment services as well as their employment outcomes. Such data are essential to carry out robust analysis. The scale and richness of data available for this evaluation allows for precision in the estimates of impact and analysis of specific cohorts of interest, typically not available in survey datasets. Nevertheless, a number of data limitations are worth highlighting. This section outlines improvements that could be made to the administrative data, pointing to some immediate opportunities and highlighting longer term development opportunities.

2.3.1. An analytical data framework that maximises existing administrative data will unlock further analytical potential

At present, the DSP does not have a longitudinal dataset for analytical purposes that enables the association of interventions and outcomes within one database. The Jobseekers Longitudinal Dataset (JLD), which collated unemployment, training and employment episodes has not been maintained since 2018. A longitudinal dataset that can be used for analytical purposes is a current development priority. Work has commenced on the Work and Welfare Longitudinal Database (WWLD) but further work will be required to develop and maintain it.

With such a dataset, high-quality analysis can be conducted in a way that enables tracking outcomes across different programmes, analysing characteristics of shorter or longer unemployment spells and attributing causal impact to the ALMPs and employment services DSP refers jobseekers to.

With access to large-scale administrative data structured for analytical purposes, the PES could provide advice to jobseekers based on the employment trajectories of people with similar characteristics. Also, analysis of the outcomes of PES programmes will lead to a greater understanding of the suite of interventions on offer and which cohorts would benefit most from particular programmes at specific points in the economic cycle. This applies not just to measuring the net impact of one programme but also to identifying the sequence of interventions that is most effective in improving jobseekers' employment prospects.

2.3.2. Data quality can be enhanced across a number of dimensions

Due to the number of different data sources and the nature of administrative data, some variation in data quality naturally exists. Most notably, in a limited number of cases, the data imply overlapping episodes suggesting either the start or end dates of eligibility, or the start and end of participation in ALMPs, are imperfectly recorded. At the point of data extraction, these conflicts are adjudicated on an ad hoc basis taking account of organisational knowledge and data reliability. A longitudinal dataset that can be used for analytical purposes, as just described, could provide some more universally applicable resolution of this kind of conflict drawing on more sources to get closer to the most accurate representation of what happened. Aside from overall comprehensiveness and coherence, the incompleteness of some variables provided further challenges. Some gaps in the data which would have enriched the evaluation include:

- PEX coverage is limited and often incomplete. The launch of Tús coincided with the rollout of PEX in 2012. 82% of Tús participants have a PEX record while only 52% of those in the eligible population who never do Tús have a PEX record. One of the main reasons for the absence of PEX records is many of those in the eligible population for Tús have claim start dates long before PEX was introduced. Where a PEX record does exist there are cases where no PEX score is recorded or answers to some questions are missing. The incomplete nature of the PEX dataset makes any analysis of this component challenging notwithstanding the fact that it is a variable that would be useful to include in the analysis.
- The DSP scheme data does not contain Jobseeker Scheme information - The dataset containing information on DSPs schemes (e.g., SPC, Carer's Allowance, Rent Supplement) for the eligible population is from a separate source to the information on jobseeker schemes (JA, JB, etc.). Again, a unified database that collates claim information from a range of schemes and adjudicates on inconsistencies in data would be of considerable analytical benefit.
- Historic DSP Payment records are limited. The data set detailing information on welfare payments made to the eligible population begins in 2010. This makes it difficult to factor in an individual's

reliance on welfare payments (or calculate the ratio as a percentage of total income) especially for those who participated on Tús or became eligible to do so during the early part of the analysis period (2011-2013).

2.3.3. Additional enhancements from external or future data sources will deepen capacity

The Government strategy *Pathways to Work 2021-2025* commits to increasing the caseload capacity of the PES and investing in digitalising PES to maximise its reach through blended service delivery (online and in person). This will play a part in ensuring labour force resilience and mobility in the Irish labour market. The following data enhancements could support this ambition by improving the evidence base on PES interventions.

Like many countries, Ireland uses a profiling tool but it involves questionnaires rather than administrative data on benefit receipt and does not take account of individuals' behaviour as they interact with employment services. DSP has ambitious plans to enhance digital engagement so it will be critical to capture data on digital interactions. Analysing how long jobseekers spend logged into services how they search through menus and whether they apply for training and jobs can provide information on their engagement. It may then be possible to use this information to categorise jobseekers and provide them with different levels of service as a result. This may be of direct use in the JobsIreland online platform to facilitate innovation in the matching of employer vacancies or placements and jobseekers shortening both unemployment duration and employer recruitment timelines.

Second the administrative employment data in Ireland do not contain information on hours worked. This is worth bearing in mind when interpreting the relationship between weeks of work and annual earnings. For example, if participating in an ALMP increases the probability that an individual will become employed on a part-time rather than full-time basis the annual earnings and weeks of employment will not reflect actual hours worked; hourly wages are in reality higher than suggested by the observed annual earnings.

Incorporating information on hours worked into the administrative employment data in Ireland could yield a signification improvement to analytical capacity. This is particularly true where participating in an ALMP may affect the probability that an individual will become employed on a part-time rather than full-time basis or as is the case with CE and Tús the placement is on a part-time basis (19.5 hours) and progression should encompass a measure of whether hours of work have increased.

Third education data provides useful information when controlling for observable characteristics that are critical to labour market outcomes. There are no administrative educational data available to inform the traditional associations between levels of education and the return to education or the building of skills and qualification through training to enhance employment prospects.

Even where data are available where the Probability of Exit (PEX) score has been calculated this is based on a questionnaire administered by DSP cases officers ("Which of the following categories best describes the highest level of education you have completed?") and may suffer from the biases and errors that affect survey-based data. For many of the people eligible for CE and Tús in the period analysed here the roll-out of PEX happened long after the beginning of the unemployment spell. A more straightforward and efficient approach is to use administrative data on educational attainment.

As a practical example the CE programme has training as a core element and data on training cover all of the training activity undertaken in the CE placement. At present it is not possible to establish a baseline of education levels (perhaps corresponding to an international standard classification) for all of the eligible population. This would enable analysis on whether training on CE was helping to address jobseekers' existing skills gaps.

With appropriate security and data privacy concerns in mind the use of the administrative education datasets to provide this information about education and training could be a useful addition to DSP's data on labour market status and interventions. A reduction in the number of questions addressed to the jobseeker on the PEX questionnaire is taking place following an ESRI review of PEX and administrative data on education would further reduce this burden.

At present there are no linked health or justice data available to examine the wider impacts of ALMPs and how the participation in an ALMP affects usage of health services or criminal acts. It is recognised that while these data may provide a more rounded perspective on outcomes or intermediate stages of progression consideration needs to be given to whether the benefits outweigh the risks associated with sensitive personal data. While it may provide useful context or secondary outcomes ultimately the range of ALMPs on offer have a primary purpose of moving jobseekers into employment.

2.3.4. Existing data can be more carefully managed

More closely related to the programme management is the practice of random selection of prospective Tús participants. The process of random selection and follow-up generates source data that are not collected on the main DSP data system but at a divisional level (DSP geographic areas of operation) with perhaps multiple inputs in each division.

The variation in the data on how people are referred to Tús and any subsequent actions appears to hinge on data recording and data collation practices in different divisions. Some divisions return complete data that documents the full journey from selection to commencement; while there is little doubt other divisions followed this procedure (as the distribution of Tús participants is not notably different) the fact that they have not returned the same scale of data documenting the journey hampers the analysis. Changes in data collection practices since the analysis period have greatly improved the data quality.

A more consistent set of practices relating to data entry and processes (perhaps hosted centrally rather than within each division) would result in a more consistent dataset for the management and analysis of the scheme. Better recording of processes at the referral stage could also inform analysis of the relationship between referral to Tús and subsequent claim closures.

These are just some of the ways the analysis could be enriched with additional databases and ways to make better use of administrative data in the future. Along with data sources there is a need for investment and resources to maximise the power of administrative data for analytical purposes.

2.3.5. Existing data should be better documented

Metadata provide source and variable information across datasets. As the dataset used in this evaluation was generated specifically for this purpose, it does not exist as part of a metadata catalogue that provides an overview of the datasets available for analytical purposes as well as their key characteristics. Ideally, if a large-scale longitudinal database is developed for analytical purposes, it should incorporate documentation, with a metadata entry for every table.

This metadata should:

- Define the sample – generally each metadata file should provide information on the sample of individuals eligible for inclusion into the dataset.
- Define timeframe – information is given on the coverage period of the data, detailing the range from earliest to latest data point. This information should detail how particular features have changed over the course of the time frame covered by the data.
- Provide information on the robustness of the variables. This is an important contextual aspect to guide an analyst on the suitability of particular variables for use in analysis. For instance, variables often have missing values but a well-considered metadata document can point to the reasons

underlying the absence of a value. In some cases, missing data may be interpreted as zero or a non-event and in other cases a missing value might indicate something other than a zero value, such as the individual in question belonging to a different population for whom it is not possible to obtain a value for them or a period of time prior to the recoding of that variable.

Provide a hierarchical rationale that allows the categorisation of an individual into a particular status, taking into account the existence of conflicting data points. The hierarchy enables a reasonable assessment to be made explicit depending on the quality of the data sources and the timeliness of updates

3 Analytical Methodologies

This chapter outlines the different analytical strategies utilised in the main report across the three topics that are evaluated.

3.1. CE Evaluation – nearest-neighbour matching

In the absence of a randomised trial to conduct policy evaluation, there are several analytical methodologies that can provide unbiased estimates of programme effects, not only the matching technique utilised for the CE evaluation in this report. For a more in-depth discussion of these different methods, a good summary can be found in OECD (2020[5]). This section outlines how participants make a choice to start CE and how this guides the analytical technique chosen for the evaluation. It describes the motivation for choosing the baseline technique and outlines the assumptions that it requires in order to produce reliable impact estimates. Alternatives to the baseline methodology chosen are also described and their advantages and disadvantages discussed.

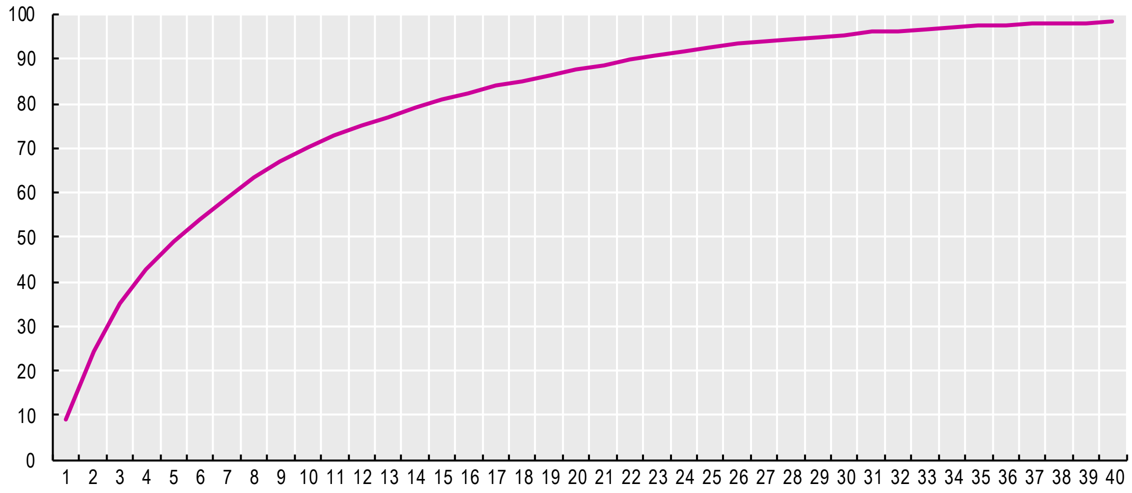
3.1.1. CE is a long-running programme with self-selected participants

The analytical technique chosen to evaluate CE is guided by implementation of the policy. CE has been running since 1994, to provide labour market integration for long-term unemployed jobseekers. It was not possible to utilise trial evidence, whereby entry to the programme would be randomised and strictly controlled, which means data on live-running programme participation for individuals that have entered a CE scheme are used. Therefore, a quasi-experimental approach needs to be adopted. These attempt to analyse participants and non-participants as if they were selected at random, by removing any other differences between the comparison of participants and non-participants.

The precise quasi-experimental technique chosen depends on the specific nature of how participants are selected from the pool of eligible individuals. Individuals are not randomly selected (unlike Tús for example) and can apply for CE positions at any point in their unemployment spell post eligibility. As Figure 3.1 demonstrates, individuals start CE at varying points after their eligibility. Whilst 10% of individuals start in their first quarter of eligibility, by the end of one full year of eligibility only 43% have commenced CE by this stage, 63% after two years and 75% after three years.

Figure 3.1. CE participation begins at very different unemployment durations for individuals

Cumulative distribution of CE starts by quarterly unemployment duration



Note: CE starts relate only to an individuals' first identified CE spell.
Source: Calculations based on Department of Social Protection (DSP) data.

The fact that individuals can start CE after such varying lengths of time in unemployment means there is a greater potential for differences in CE participants, both compared to one another and to the broader eligible pool of non-participants. In addition to the difference in the starting points of CE participants in their unemployment claim, the fact that individuals decide for themselves whether or not they wish to participate means that participants can differ from eligible non-participants in other ways. For example, younger people may prefer to participate relative to older people or there may be more female participants than male.

3.1.2. Dynamic propensity-score matching is used to estimate programme effects

To address these issues, this report uses a strategy of dynamic propensity score matching to ensure that individuals who participate in CE are compared to only a group of individuals who are similar to them. This approach aims to ensure that any differences that exist in outcomes between the two groups are the result only of participation in CE. Where programmes can take place at varying points in a person's unemployment spell, it has been shown that it is important to take account of differences in individuals dynamically to ensure that compositional differences between participants and non-participants are factored in (for examples see Sianesi (2008,[6]) (2004[7])). Standard "static" evaluations, which compare participants to individuals that never participate, may induce biased estimates in situations where programme entry can occur at various points in time. For example, by comparing CE participants to only those individuals that never participate in CE, an analysis may incorrectly limit the comparison group to only those individuals that did not participate because they quickly found a job. Therefore, this analysis looks at quarterly cohorts of jobseekers and compares CE participants to all eligible participants, including those that may start CE at a later point in time.

Due to the nature of the programme and its eligibility only for individuals with more than 12 months of unemployment, a more relaxed specification of dynamic matching is implemented and matching is not conducted solely on individuals with an unemployment duration at least as long as the CE participants (in Sianesi's (2004[8]) specification, participants were matched with individuals with the exact same duration of unemployment). Instead, duration enters into the regression specification for the probit regression that is used to estimate participation probabilities, so that it allows individuals of different durations to be matched to one another, if after adjusting for the impact of duration on participation, there exists an individual with a set of characteristics that better represents the CE participant in question.

The dynamic specification of the CE evaluation compares the labour market outcomes of those who begin the ALMP in that quarter with those who are still "waiting" for support from an ALMP measure or for some other way out of unemployment. Using quarterly cohorts of individuals ensure that potential effects of economic cycles are automatically accounted for in the analysis. CE participants are compared only to those non-participants that were eligible in the same calendar quarter. Estimates for quarterly cohorts are aggregated for the time horizon of interest (t, the amount of time elapsed since the start of the ALMP measure, when labour market outcomes are measured). The *potential* labour market outcomes (such as employment or earnings) for an individual (i) can be written Y_{iqt}^{d}, where $d = 1$ under treatment and $d = 0$ otherwise.[2] The average treatment effect on the treated ($D_{im} = 1$) for each t can then be written:

$$\gamma_t = E\left[Y_{iqt}^{1} \middle| D_{iq} = 1\right] - E\left[Y_{iqt}^{0} | D_{iq} = 1\right]$$

In this framework:

- The treatment group comprises those individuals who *begin* treatment in quarter q.
- The potential comparison group comprises individuals who were still unemployed in period q, but who did not enter an ALMP (including those that went on to start training in future periods).

The dynamic selection-on-observables approach therefore recasts the simple comparison between participants and non-participants – which would lead to biased estimates of programmes' treatment effects – as a series of more reasonable comparisons between individuals with similar experiences of unemployment at that point. The approach also tackles the issue of multiple treatments by focussing on the *first* ALMP measure that individuals receive. It treats any subsequent ALMPs as part of individuals' labour market outcomes. Thus, if an individual stays unemployed so that they can enter another ALMP (or enters another ALMP because they have remained unemployed), this is treated as information about outcomes rather than information about subsequent treatments.

A wide array of variables is used in the propensity score calculations. These variables can be grouped into the following categories:

- **Prior labour market histories:** previous known occupation, previous annual weeks worked and annual earnings for the three years prior to CE eligibility.
- **Prior benefit histories:** length of time in receipt of unemployment benefits, family benefits, caring and disability benefits, rent supplements for each of the three years prior to CE eligibility. Total prior lifetime unemployment benefit spells and duration.
- **Demographic characteristics:** age, gender, marital status, presence of children in family and location
- **[For the PEX sub-group]:** Education, willingness to move for work, native language, self-assessed reading and writing ability, motivation for job search.

These extensive set of covariates, which include prior labour market and unemployment histories have been shown by several studies to mitigate the effects of unobserved characteristics in constructing comparison groups (see (Heckman et al., 1998[9]); (Lechner and Wunsch, 2013[10])) Caliendo, Mahlstedt and Mitnik (2017[11]) provide evidence that usually unobserved aspects such as personality traits, social networks and life satisfaction played an important role in determining selection into ALMP treatment in

[2] To facilitate the description of the identification approach, the exposition in this section describes an approach which does not employ difference-in-differences to account for unobserved heterogeneity. A similar framework can be adapted to the approach difference-in-differences approach used in the (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) impact evaluation.

Germany, but that conditioning on detailed labour market histories can implicitly capture much of the information contained in the usually unobserved variables. This report makes a similar finding. It has also been able to make use of a detailed set of questions utilised for risk-scoring of unemployed individuals (PEX questions) that allow testing of potential unobserved variable bias in the context of CE participation. In particular extra information on education, reading, writing and language abilities, and motivational questions on jobsearch (such as the willingness to move for work). The results of which are presented later in this chapter.

Previous research has also shown that experimental and quasi-experimental approaches (such as the matching technique utilised here) do not systematically vary in terms of finding "less positive" or "less significant" results (Card, Kluve and Weber, 2018[12]). This, coupled with the knowledge that studies have shown the efficacy of detailed administrative data for unbiased programme estimates, mean that the array of administrative data in this report and the additional ability to incorporate more detailed socio-economic and attitudinal questions on a sub-sample of participants, provide a high degree of confidence that estimated impacts provide a good reflection of how CE helps individuals to progress in the labour market.

3.1.3. The nearest-neighbour matching strategy provides desirable bias properties

The literature on counterfactual impact evaluations provides many different plausible estimators that this report could use to derive programme effects. Due to the large sample sizes available for this report- with almost 30 000 participants and 620 000 unique eligible non-participants, the (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) impact evaluation uses a nearest-neighbour matching estimator. Huber et al. (2013[13]) detail that this exhibits the smallest bias for all sample sizes. Bias is important because it refers to the extent to which an estimate may not reflect (be bigger or smaller than) the "real" outcome it is trying to measure. Using the most similar controls (i.e. the individual that resembles the participant most closely out of all of the eligible non-participants) reduces bias and exhibits robustness to propensity score misspecification because it remains consistent if the empirical propensity score specification is a monotone transformation of the true model. Put differently, so that if the "true" model were different, as long as the order of individuals was the same, in terms of their probabilities, it would not affect the matching procedure. Its disadvantage is that it is inefficient- by utilising fewer observations than are possible in the data, the variance of the estimates is higher (it is more difficult to be certain that impacts exist). However, with large sample sizes, such as those available in this report, the superior properties in terms of bias are relatively more important, with variances tending to zero asymptotically.

Many other types of estimators could have been employed for estimating the average treatment effect. These include parametric estimators (such as doubly-robust estimators), inverse probability weighting estimators, matching estimators, and kernel matching estimators. These estimators all have benefits and drawbacks, and studies have not ambiguously agreed on which estimator is best. For example, Frölich (2004[14]) finds that a version of kernel-matching based on local regressions with finite sample adjustments performs best in Monte Carlo simulations with up to 1 600 observations. By contrast, Busso et al. (2014[15]) show inverse probability weighting using normalized weights exhibits the best properties in finite samples where overlap is good, via the use of Monte Carlo simulations and empirical data-generating processes. They find that in instances of data-generating processes with poor overlap, bias-corrected matching with a fixed number of neighbours is the most effective. Huber et al. (2013[13]) perform an empirical Monte Carlo study on a large number of estimators measuring average treatment effects on the treated and find that no estimator is superior in all designs and for all outcomes and that bias-adjusted radius matching estimators perform best on average. The relative performance of estimators arguably depends strongly on features of the data-generating process (Busso, DiNardo and McCrary, 2014[15]) which is unknown to the empirical researcher in practice.

Matching relies on the assumption of conditional independence

A fundamental assumption in the propensity score matching used in this report is that all differences between CE participants and eligible non-participants that affect both their participation and their future outcomes are captured in their observed characteristics (that is in the administrative data variables the report use). This means that after conditioning on observed covariates selection into the treatment can be considered random (Imbens, 2000[16]). In the case that selection into programme is governed not only by observable but also *un*observable individual characteristics that are correlated with the potential outcomes, then propensity score matching provides biased estimates of treatment effects. To address this issue, previous studies have combined propensity score matching with a difference-in-difference approach (Heckman et al. (1998[9]), Smith and Todd (2005[17])). This approach compares intertemporal changes in outcomes between participants with changes in outcomes for the comparison group, with changes measured relative to a pre-programme benchmark period:

$$\Delta_{im}^{\tau} = \left(y_{im,t=\tau}^{1} - y_{im,t=-\tau}^{1}\right) - \left(y_{j,t=\tau}^{0} - y_{j,t=-\tau}^{0}\right)$$

where y is the actual outcome of interest, $y_{im,t=\tau}^{1}$ and $y_{im,t=\tau}^{0}$ are post-unemployment outcomes for participants for controls, respectively, and $y_{im,t=-\tau}^{1}$ and $y_{im,t=-\tau}^{0}$ are their pre-programme outcomes. By controlling for time-invariant unobserved heterogeneity, this specification ameliorates the bias resulting from unobserved differences between participants and non-participants.

Placebo tests on past outcomes, statistics on overall matching quality of the model and the additional analysis conducted using PEX questions all provide confidence that the nearest-neighbour matching has accounted for both observed and unobserved differences (particularly as the pre-period outcomes demonstrate good balance around zero). As an extra pre-caution against any omitted variable bias due to unobserved bias, sensitivity analysis is conducted which implements the matched difference-in-difference approach described above, the results of which are presented below.

Alternative identification strategies based on policy design were not feasible

One alternative identification strategy is to use programme eligibility rules to compare those individuals either side of eligibility thresholds. CE eligibility is fairly broad and is determined mainly by virtue of being long-term unemployed, however there is policy design related to age and maximum length of participation that were considered as part of the evaluation in this report. It was decided that these policy thresholds were not suitable to use for evaluation purposes.

Conducting a regression discontinuity around lifetime participation limits was discounted. During the evaluation period, people under 55 could participate in CE for a maximum of three years, and people over 55 for a maximum of six years. This approach would compare two individuals close to the cut-off on participation, one who was just over and one who was just under. The individual who was just under the limit and then participated in the programme is then compared to the person who was just ineligible for CE, on the basis that the only difference between these individuals was the participation in CE.

However, this approach was dismissed because of its practical limitations. An individual could always choose to wait to “age into” a new higher participation limit. Those ineligible 54-year-olds soon become eligible 55-year-olds (usually a research design based around age relies on eligibility being excluded as individuals get older- think of programmes for young individuals).

A final practical limit to this approach was that the lifetime limits were such (six years or three years of participation) that it limits both the number of individuals reaching that point and the ability of those estimates to provide information for different CE claimants. For example, determining the impact of an individual completing a 4th episode of CE might not provide much in the way of translatable impacts to an individual who only completes one episode. For these reasons, a research design exploiting CE eligibility limits was discounted.

3.2. Tus evaluation – inverse probability weighting

Selection and referral to Tús occurs on an ongoing basis. The analytical approach is to assess eligibility for the treatment and control population on a quarterly basis, capturing movements in and out of eligibility throughout the analysis period.

The analysis of Tús applies the dynamic selection-on-observables approach to compare people who commence Tús in a quarter to people who are eligible but have not commenced Tús. Each quarter is treated entirely separately for the purposes of modelling participation in Tús, therefore capturing the variability in labour market prospects facing the long-term unemployed over the period 2011 to 2018.

Inverse Probability Weighting (IPW) is the method applied in the Tús evaluation. Rather than selecting control units to be in the comparison between treatment and control, inverse probability weighting retains all units and weights the control group to resemble the treatment observations.

In theory, the random selection from the Live Register, Ireland's register of jobseeker claims, fulfils the conditions for a randomised control trial (RCT). In a RCT, the treatment variable is disconnected from variables that influence the outcomes. Any differences in observed outcomes can be solely attributed to the treatment. If this were the case for Tús, then evaluating the causal impact of Tús on participants should be free of self-selection or administrative selection bias.

With Tús however, opportunity exists for both selection and administrative bias. There is scope for self-selection as not all of those who are referred to an initial Tús interview will attend/begin a placement. In the absence of a complete dataset on Tús referrals it is not possible to confirm that the referral process is completely random. In addition, Tús referrals are administered regionally which results in variation in reporting procedures, meaning it is not possible to completely rule out the potential for administrative bias. While initial selection may be random, commencement on the programme may reflect negative or positive selection. After selection, willingness to take up a Tús placement might be due to greater motivation (in which case participants are positively selected with regard to the control group) or greater need of the placement (in which case participants are negatively selected with regard to the control group).

In summary, some degree of adjustment is required to render the control group similar to the treatment group across key dimensions. With the range of administrative data variables available, observed differences between the two groups can be corrected for, re-affirming the approach outlined below.

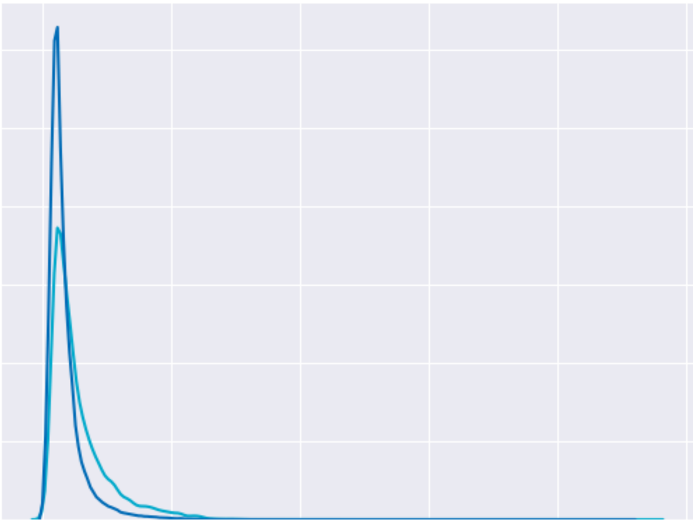
Modelling participation in Tús starts by comparing participants in each quarter between Q3 2011 and Q4 2018 to the set of eligible non-participants in the same quarter where beginning Tús in a given quarter is a binary outcome (0,1). Propensity scores are generated, measuring the person's likelihood of starting Tús in a given quarter. The score is a continuous measure between 0 and 1, taking into account the individual's employment history through variables such as previous earnings and periods of unemployment and demographic characteristics like age, sex and nationality. The model uses propensity scores to generate weights which are then applied to construct a control group similar to the treatment in each quarter.

After weighting, the second step is the outcome estimation. Weighted Least Squares (WLS) regression is applied to measure labour market outcomes such as earnings from employment, weeks spent in employment, weeks spent in receipt of welfare payments and weeks spent on the unemployment register. This step takes a "doubly robust" approach where the second stage regression also controls for a range of both demographic and employment-related characteristics.

3.2.1. Verification of methodology Positivity check

In the case of random referral to Tús, it is perhaps unsurprising that there is an overlap in the treatment probability distributions. As this overlap holds across the entire population, rather than just partially, there is no requirement to restrict analysis (for example, only to specific age cohorts, or only to men). The assumption is necessary to ensure valid comparisons can be made between different groups without introducing biases. It also has implications for generalising results. Figure 3.2 shows the overlap between treated and non-treated groups on the basis of the probability of participating in Tús.

Figure 3.2. Positivity check - distribution of propensity by treatment status



Source: Calculations based on Department of Social Protection (DSP) data.

In using the full eligible population, the objective is to construct a weighted dataset such that the treatment and control groups are similar on the basis of pre-treatment values. Subsequently, the analytical output is an assessment of the variance in post-treatment values. As illustrated in Chapter 4, and as is typical in these scenarios, the data start in an imbalanced state between treatment and control groups.

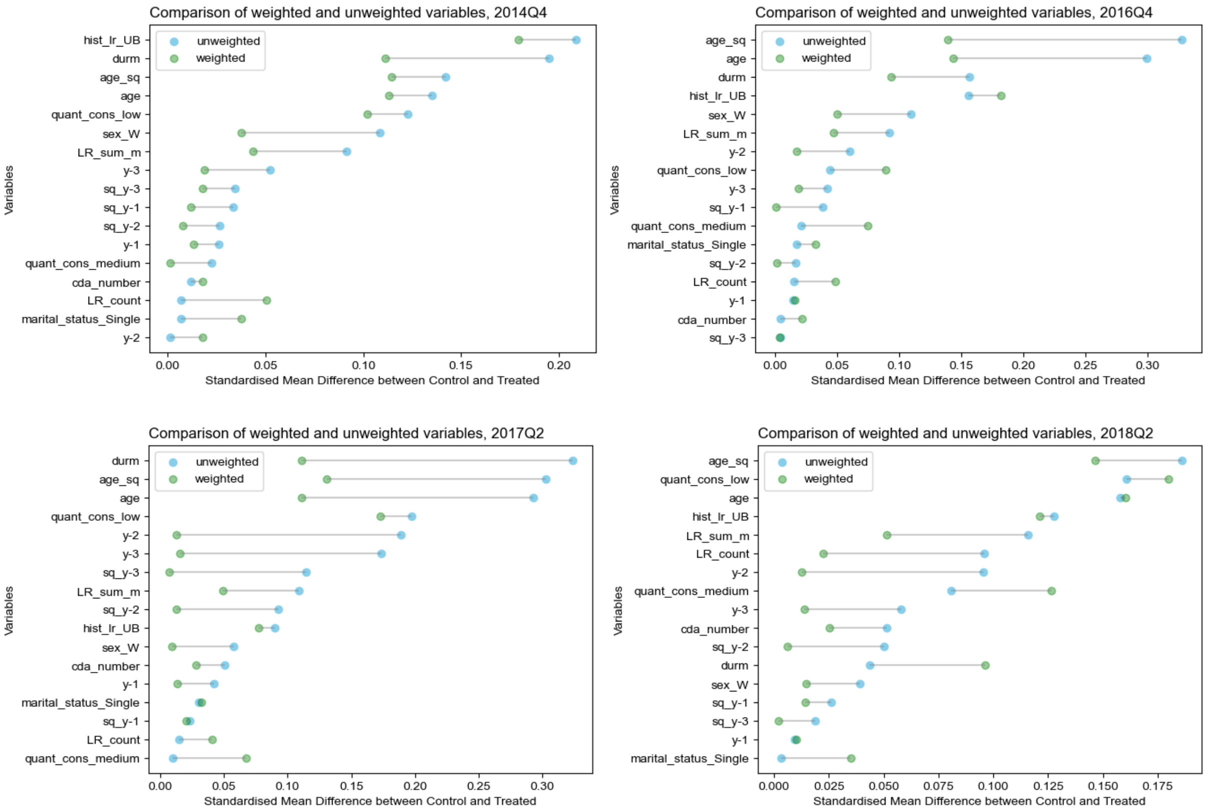
The concern is in cases where there are relatively large differences between the treatment and control cases along the covariates of interest, with the ideal scenario being that treatment and control groups have similar mean values on relevant covariates.

The re-weighting that is conducted to account for differences in individuals is successful in ensuring that the treatment and control groups are similar to each other. The effectiveness of the re-weighting using propensity scores can be examined by looking at how similar control and treatment groups are in the periods before programme eligibility across key variables. For example, by looking at the similarity of values for past earnings and unemployment histories between the treatment and control groups confirms that the re-weighting procedures are successful.

This has two implications. Firstly, it shows that despite random selection for referrals, there is non-random selection into participation, such that participants (the treated) differ systematically from non-participants (the controls). Second, it shows that the evaluation methodology is successfully able to adjust for this non-random selection, such that estimates for Tús impacts can be considered "as good as random" and show causal impacts of Tús participation.

The plots below show the balance of covariate distribution between treatment and control groups. The standardized mean difference is independent of the unit of measurement and therefore permits a comparison between variables that have very different units of measurement. The plots below present a sample of quarters in the analysis period and illustrate the large reduction in the weighted values compared to the unweighted values.

Figure 3.3. Weighting reduces the bias between participants and non-participant in Tús



Source: Analysis of Department of Social Protection administrative data

3.3. ALMP Sequence Analysis

This section describes how the administrative data for the sequence analysis are organised. It provides the underlying rationale describing how the choices on data aggregation are made. It then goes on to consider what extensions may be made on sequence analysis, to investigate sequences in a counterfactual methodology, such has been done in the narrower CE and Tús evaluations. This could provide further information on whether one sequence is preferable over another in terms of individuals future outcomes, work that is beyond the scope of this current report.

3.3.1. The time spent in each scheme matters

Spending a month or a year in any of the observable schemes presumably makes a difference. On the one hand if – for instance – the scheme is paralleled with training, the longer the beneficiaries stay, the more training they receive, what presumably enhances their chances to get a job. On the other hand, the longer they stay outside of the labour market, the less likely may become to get in again. Effects are therefore a priori undecided. Duration analysis concerns all the methods designed to ascertain these effects. Applications in the social sciences are pervasive: what is the effect of staying longer in unemployment? Does one lose employment opportunities due to skill obsolescence, or get better and more stable ones as he disregards the bad ones? And also: what is the effect of staying married longer? Does the probability to split increase or decrease when people know each other better? Duration analysis has to overcome a main problem, labelled *dynamic selection*: the sample under scrutiny typically changes in a non-random way as time goes by. To go back to the example of unemployed people who look for a job, there are at least two effects going on: first, as described above, the longer one stays unemployed the more obsolescent her skills will become, what reduces employment opportunities; second, as time goes by, the best workers leave the sample as they find a job sooner on average, so the sample remains with less skilled workers, who enjoy poorer chances to find a job. The latter is dynamic selection. At long unemployment durations, hence, we may hence observe that the probability to find a job decreases for two reasons: skill obsolescence – which concerns all workers irrespective of their innate ability – and dynamic selection. Duration analysis aims to get rid of the latter to close off the former, also called (true) *duration dependence*. In its proper sense, duration dependence means that prospective duration in the current state depends on the time elapsed in the state itself. Negative (positive) duration dependence implies that the longer one stays in the observed state (e.g., unemployment) the less (more) likely will be to leave it.

Duration analysis is used in Chapter 5 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1])to study the probability – conditional on observables and on time elapsed since entry into the relevant state – that an individual eligible to CE or Tús, or – separately for each state – receiving CE or Tús, may move to another of the observed states. Typical concerns of duration models are:

1. *Right-censoring*. Data – including the sample used in Chapter 5 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) – necessarily includes observations referring to individuals who have not yet experienced an event bringing the duration under scrutiny to an end, as this event may well come in a future as yet unobserved. Sampling these individuals out is not a solution, as it would be tantamount to retain in the data only the shortest durations, i.e., selecting on the dependent variable, what would lead to biased estimates. It is usually assumed that right-censoring is not informative, i.e., that it is random with respect to the durations themselves. This assumption typically holds when right-censoring is due to an exogenous event, e.g., the date of last data update. This is indeed the case, as availability of data limits the observed period to December 2019.
2. *Left-truncation*. Analogously, when a series of data becomes available, a subset of durations is already ongoing, which implies that an unknown portion of such durations remains unobserved. In this case what is called flow sampling (i.e., retaining in the data the durations starting from one date onwards) is a suitable solution, as – provided right-censoring is dealt with as described above – durations are not artificially shortened. This is implemented in Chapter 5 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), where all eligibility periods starting between 2012 and 2015 are selected for the analyses.
3. *Dynamic selection*. Duration models – through conditioning on elapsed time into the current state – aim at controlling for dynamic selection, and hence to isolate true duration dependence.

In (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), duration models are estimated for CE/Tús eligibility, CE and Tús periods, assuming that transitions out of the current state occur discretely at monthly intervals, consistently with the structure of the data. While the needed assumption to make this approach viable – i.e., that the distribution of the transitions within each month is uniform, see Jenkins (Survival Analysis, 2005) – appears acceptable, the estimation of duration models in discrete time has at least three advantages. First, once the data are rearranged in a person-period structure – as conducted in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) and recalled above – estimation boils down to a (computationally simpler) multinomial logit model in which the dependent variable takes the value *0* for the entire duration under scrutiny but the very last month, when it either i) keeps being *0* if the duration is right-censored, or ii) takes a different value for each mutually exclusive event that brings the duration to an end. Second, this setting allows for a semi-parametric specification of the duration process through a piecewise constant spline that do not impose any shape to the duration dependence structure. Third, as Dolton and

Van Der Klaauw (The turnover of teachers: a competing risk explanation, 1999) suggest, such flexible functional forms for the duration process are also more robust to the presence of unobserved heterogeneity.

Formally, and taking the duration of eligibility spells as an example, models of the following type are estimated:

$$\Pr\{K_{i\tau} = k\} = \Lambda\left\{X_{it} + \sum_{q=1}^{7} Q_{iq\tau} + \sum_{y=2012}^{2019} Y_t + \sum_{s=1}^{4} S_{st} + \varepsilon_{it}\right\}$$

where:

4. subscript *i* stands for the eligible individual, τ for the individual time (measured in months) since entry into eligibility, *t* for the calendar time. Calendar and individual time can be identified separately. Indeed, as we observe individuals entering eligibility over several years, one can well be – e.g., – in the first month of eligibility during March 2012, or September 2015.
5. $K_{i\tau}$is the set of mutually exclusive and exhaustive states, where *k* = 0 if ongoing or censored, 1 if moving to CE, 2 to Tús, 3 to Employment with Support (EWS), 4 to Employment without Support (EWoS), 5 to any other state; the existence of multiple exits makes ours a competing-risk duration model (see below for a more precise definition of EWS and EWoS).
6. X_{it} is a set of control variables including age, gender (through a dummy for females), marital status (through a dummy for married individuals), nationality (through a dummy for Irish nationals), and years accrued in the Live Register (discretely with categories: no time accrued, until one year, one to two years, two to five years, more than five years). All controls are observed at entry, but age that varies over the (elapsed) time since eligibility.
7. $Q_{iq\tau}$ is a set of seven four-month-intervals-specific time dummies measuring time elapsed since entry into eligibility (the first six dummies are specific to the first six intervals, while the seventh refers to more than two years since eligibility). This is the way our competing-risk model controls for dynamic selection in discrete time with a functionally flexible approach.
8. Y_t is a set of dummies for calendar years (from 2012 to 2019) and is meant to control for business cycle effects, while S_{st} is a set of four calendar months, aimed at controlling for seasonality.

With this kind of specification - and assuming that the set of controls is sufficiently large to minimize the room for confounding effects, $Q_{iq\tau}$ identify the duration dependence profile, while the set of controls X_{it}, Y_t and S_{st} represent vertical shifter that may raise or lower the time profile in a parallel way. To put it differently, control variables do not affect the shape of the duration dependence profile, but only its vertical position.

In their meta-analysis of ALMPs from 97 different studies, Card, Kluve and Weber (Active labor market policy evaluation: a meta-analysis, 2010) find that 20.6% of ALMPs have a duration of up to four months, 35.2% of between five and nine months, and 18.1% of over nine months, with a remaining 26.1% of cases for which duration is unknown or mixed. Most programmes are therefore short, and even shorter in Anglo-Saxon countries. This is however not the case for Ireland, where Tús and CE in particular have much longer durations. Duration models are used in 36.2% of the cases studied by Card and colleagues, mostly to analyse unemployment duration until a job is found (24.6%) or until any type of event occurs (8.5%). Empirical meta-analysis by Card, Kluve and Weber (2010) suggest that duration models tend to show positive short-term impacts than evaluation based on direct labour market outcomes, and that longer programmes are not more effective than shorter ones.

3.3.2. The order and timing of ALMPs may have important impacts on outcomes from programmes

Beyond the length of stay in a single programme, combinations and timings of ALMPs are important for PES to consider. Whether to refer a jobseeker to training before or after a period of job search support is not a trivial matter and may affect that individual's future pathway in the labour market. In this sense, it is important to know not only whether, for example, training works in isolation (against not participating in training) but also whether that training has different effects in combination with other programmes (for example, whether it is more effective prior to job search support than after it). Accumulating evidence on these different programme pathways can help PES to make better programme allocation decisions for their clients and provide better value for money to the taxpayer.

However, despite its potential significance, this type of causal programme sequence analysis is not something that is prevalent in the literature on ALMPs. For example, in their meta-analysis of ALMPs, Card, Kluve and Weber (2018[18]) incorporate estimates from 207 impact evaluations, covering the suite of ALMPs that are deployed by PES, but it makes no reference anywhere in the paper about the potential effect different sequences might have on potential outcomes from policies.

The existing literature on causal sequences of ALMPs is developed around work stemming from Lechner and Miquel (2001[19]), building on the earlier work of Robins (1986[20]) in epidemiology and biostatistics. The basic premise in this dynamic literature is to use participation probabilities at each stage of the sequence (the probability that an individual takes a certain course of action at a particular point), including the use of intermediate outcomes, to estimate participation in subsequent states. These estimates are used to adjust weighting for individuals in one sequence so that they are compared to similar individuals in another sequence. This means that multiple stages of a sequences are used to adjust which individuals from the respective full sequences are compared to one another. Lechner (2004[21]) proposes two estimators using inverse probability weighting and matching. These estimators are prevalent in the static impact evaluation literature and are used frequently for evaluations of individuals programmes.

The dynamic literature uses information on how previous participation affects outcomes which then affect participation in the next stage of the sequence. The use of intermediate outcomes for stages of sequences recasts the static evaluation problem as a series of steps in which outcomes following previous programmes are used to define participation probabilities. For example, it allows for the outcomes following a course of job counselling to determine the probability of that individual then participating in a training programme. Using these intermediate outcomes allowing removal of any selection bias that arises in this later period of the sequence. For example, if individuals were first enrolled in a programme of job search support and this helped individuals into work, which then affected their later decision on whether to participate in training (regardless of whether they were still in work at the particular moment of that decision), then having information on that work is important in explaining that choice. The dynamic framework allows consistent and unbiased estimation of these choices. It relies on the weak dynamic conditional independence assumption (W-CDIA). This re-casts the conditional independence assumption of the static literature, which states that outcomes are independent of programme participation, after controlling for observed confounding variables. The W-CDIA states that at each sequence stage outcomes are independent of participation having controlled for confounding variables, including intermediate outcomes variables until the beginning of that phase of the sequence.

What do the studies on sequencing conclude?

The studies that account for these dynamic casual pathways are able to demonstrate that different sequences of programmes do indeed generate different labour market outcomes for individuals. In a paper considering One-Euro jobs in Germany (a direct job creation scheme that most closely resembles the CE schemes considered in this report) Dengler (2015[22]) considers whether it is better for individuals to participate in a One-Euro Job in the first period of a two period sequence, against a counterfactual of unemployment benefit receipt, or to participate in a One-Euro Job in the second period. The paper finds that participation in the first period of unemployment generates better outcomes than participation in the second period. The paper also considers whether two consecutive periods of One-Euro Jobs generates better outcomes than two periods of unemployment benefit receipt. Here is finds that repeated participation in One-Euro Jobs produces better labour market outcomes than two periods of benefit receipt. This is relevant for the study of CE in Ireland, as repeated participation in CE is something that is commonplace for jobseekers. However, this effect is not universal. Dengler (2015[22]) splits the analysis into sub-groups by gender and whether or not the location is East or West Germany. For men in East Germany, there is evidence of ‘programme careers’ whereby participation in two periods of One-Euro Jobs sees participants with worse future labour market outcomes than counterparts who received only two periods of unemployment benefits. Hence, local labour market conditions and innate characteristics of individuals may influence to what extent these direct job creation schemes have the ability to better connect individuals with employment in the open job market.

Dengler (2019[23]) considers sequences of training for adults, looking at both timing and whether sequences of repeated training offer better outcomes than periods of unemployment benefit receipt. The results show that two periods of training produces better outcomes than two periods of unemployment benefit and also that earlier training produces better outcomes than later training. In this sense it shows that a better strategy (for the German PES) is to offer training to jobseekers earlier, than to wait for them to look for a job and only offer training to them after a period of them not having found one.

Lechner (2012[24]) provides an application of dynamic inverse probability weighting to consider ALMPs in Switzerland. The paper provides evidence that two periods of wage subsidy helps connect individuals with jobs better than two periods of unemployment benefits. This contrasts to employment programmes, where no positive or negative effect can be determined following two periods of participation, relative to remaining unemployed. This is attributed to their large lock-in period, where the typical duration would be for 6 months, meaning that individuals delay their periods of job search. A comparison of the effectiveness of two periods of wage subsidies to two periods of employment programmes, provides tentative evidence that wage subisidies generate better labour market outcomes, though there is insufficient precision in the estimates to be definitive.

Lechner and Wiehler (2013[25]) consider sequences from both a timing perspective (looking at earlier or later participation) and whether particular orders and combinations of programmes are more effective. Broadly, they find that earlier participation in ALMPs produces better outcomes than later participation. Participation in active job search or qualification and training in either the first or second trimesters of initial unemployment produces better outcomes for individuals than delaying that participation until the third trimester. Looking at sequences of programmes, they find that a qualification measure followed by active job search is better for individuals than doing the active job search before the qualification measure. However, this is contrasted to when they study the order of orientation and qualification measures, when no evidence can be found that one particular programme is better before the other. They also consider two periods of training versus only one period of training and find that two periods of training produces better long-term labour market outcomes for individuals, than one training participation in the first trimester of unemployment.

Sequencing is a complementary way to look at the problem of dynamic selection which is planned to be applied at a later stage of analysis.

4 CE Analysis Robustness Checks

This technical report considers two types of robustness checks on the main results of the CE analysis. The first centres around varying the technique utilised for the baseline analysis and the assumptions implicit within this technique. The second makes use of different data to analyse the degree of variability in the result, compared to the baseline analysis used in the main report.

For the robustness checks on techniques, a number of approaches are used. The first, most similar in nature to the original analysis, is to use a different matching algorithm to select control participants to compare to CE participants. This method provides an intuitive feeling for how sensitive the results are to particular matching algorithm used in the evaluation. As described in section 3.1, nearest neighbour matching was utilised as it has been shown to produce low bias, at the expense of larger variance as more observations are "thrown away" and the standard errors are calculated from fewer observations. For the robustness checks on matching, the report utilises Kernel matching, this has the advantage of reducing the variance (or "uncertainty") of the estimates as the whole sample is used. Alongside this, a technique based on re-weighting individuals (inverse-probability weighting, IPW) is also employed. This uses the propensity scores from the first stage participation regression (the probability that someone starts participating in CE). Rather than use these probabilities to "match" individuals with similar probabilities, it uses these scores to increase or decrease the contribution of an individual to the final outcome estimate. It increases the weight given to eligible non-participants with a higher estimated participation probability and lowers the weight for CE participants with a lower participation probability.

The final sensitivity on techniques is performed using matched difference-in-difference analysis. This uses the baseline matching specification, but compares only changes to outcomes variables relative to their level in the year three years before CE participation. To the extent that the baseline matching has already performed a good job in comparing participants and similar non-participants for these pre-participation outcomes, this differencing should have a small impact on the final estimates. Reviewing whether this is the case can provide extra information on whether the baseline specification is successful at removing any bias. It means that the assumption used in matching, the "conditional independence assumption" no longer has to hold and instead all observed and unobserved differences that are stable over time are automatically accounted for. Instead, it relies on a separate assumption that states that any differences between comparison cohorts must remain stable over time. This can be checked in a visually similar way to the conditional independence assumption on matching, by reviewing whether there are any pre-period trends (in the years prior to eligibility) in the difference in outcomes between participants and non-participants

Changes to the data sample take place along two principle dimensions- analysing annual CE cohorts, to determine how much results vary, at similar time horizons post-CE participation and by analysing a subset of the CE participants (and eligible non-participants) for whom PEX questionnaire information is available. This PEX information will allow the report to better establish the potential for omitted variable bias in the overall results, to scrutinise whether or not having more socio-demographic and motivational information on programme participants changes the magnitude and/or sign of the programme estimates.

4.1. Nearest-neighbour matching produces the most accurate evaluation results

The matching algorithm used in the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), nearest-neighbour matching with replacement, was chosen for its potential to minimise bias (bias meaning where estimates of impacts do not accurately represent the true impacts of the programme). Given the large volume of administrative used in the analysis, the trade-off that this specification brings- namely greater variance due to the use of fewer observations- is not of grave concern and the advantages more than outweigh the disadvantages.

The sensitivity analysis conducted confirms that the choice of nearest-neighbour matching does provide the specification with the least bias, relative to both inverse-probability weighting and kernel matching. The exception to this is the matched difference-in-difference technique which improves these pre-period diagnostics slightly compared to the baseline matching alone.

This can be confirmed graphically in Figure 4.2 when reviewing how well each analytical specification balances the outcome variables in the period prior to participation. If CE participants are compared to a group of non-participants that are similar to them, there should be no differences in these variables prior to the period of participation. Panels B (Annual Employment Weeks) and C (Any Employment in Year) of Figure 4.2 demonstrate where the difference-in-difference technique is able to better balance pre-participation outcomes in years -3 to -1, relative to the baseline specification. For these two outcome variables the difference-in-difference estimates are all insignificant, whereas the baseline specification is significant for each of the three years for "Annual employment weeks" and for years -3 and -1 for "Any employment in year". Despite this significance, suggesting that matching has not perfected balanced these characteristics, it does not invalid the final results, given the other diagnostics the report reviewed and the techniques that are employed. The overall model balance is good (see below), doubly robust estimates further control for any bias, and quantitatively the results of the baseline matching specification and the matched difference-in-difference specification are very similar. In addition to this, the volume of data that is analysed mean estimates can be statistically significant even where they remain economically insignificant. That is to say, that given a sufficient volume of data even tiny movements of estimates appear as statistically significant. Therefore, it is necessary to employ a degree of expert judgement when reviewing how important these statistically significant differences are in the "real world" and not only the statistical one.

Bias can also be checked by reviewing how well the matching algorithms have minimised the total differences in characteristics of individuals utilised in the propensity score specification. For example, when comparing the differences across individual characteristics of CE participants to all eligible non-participants, there is an average standardised mean bias across all of these characteristics combined of 13.3%. This overall standardised mean bias allows all relevant characteristics to be considered together, even where they have very different units of measurement. It looks at how much variation there for each individual characteristic, relative to its own size. For example, consider the following two differences. A difference in average earnings of 300 EUR on a baseline of earnings level of EUR 10 000 represents a 3% bias. A difference in weeks worked of 5, compared to an average week's worked of 50 represents a 10% bias. Standardised mean bias averages the percentages (3% and 10%) whereas a calculation that only included levels (300 and 5) would give much more weight to the earnings variable.

The baseline specification in this report reduces this overall bias to 2.2%. By contrast, uses the Kernel matching algorithm is successful only in reducing this bias to 11.3%. Caliendo and Kopeinig (2008[26]) suggest that achieving a total mean bias of between 3% and 5% after matching suggests the matching has been successful at balancing the characteristics of participants and non-participants, so the baseline nearest-neighbour specification performs well in this respect.

As the main report detailed, an extra safeguard was taken by making the results "doubly robust" by utilising regression analysis on top of the matching undertaken to ensure participants look like non-participants.

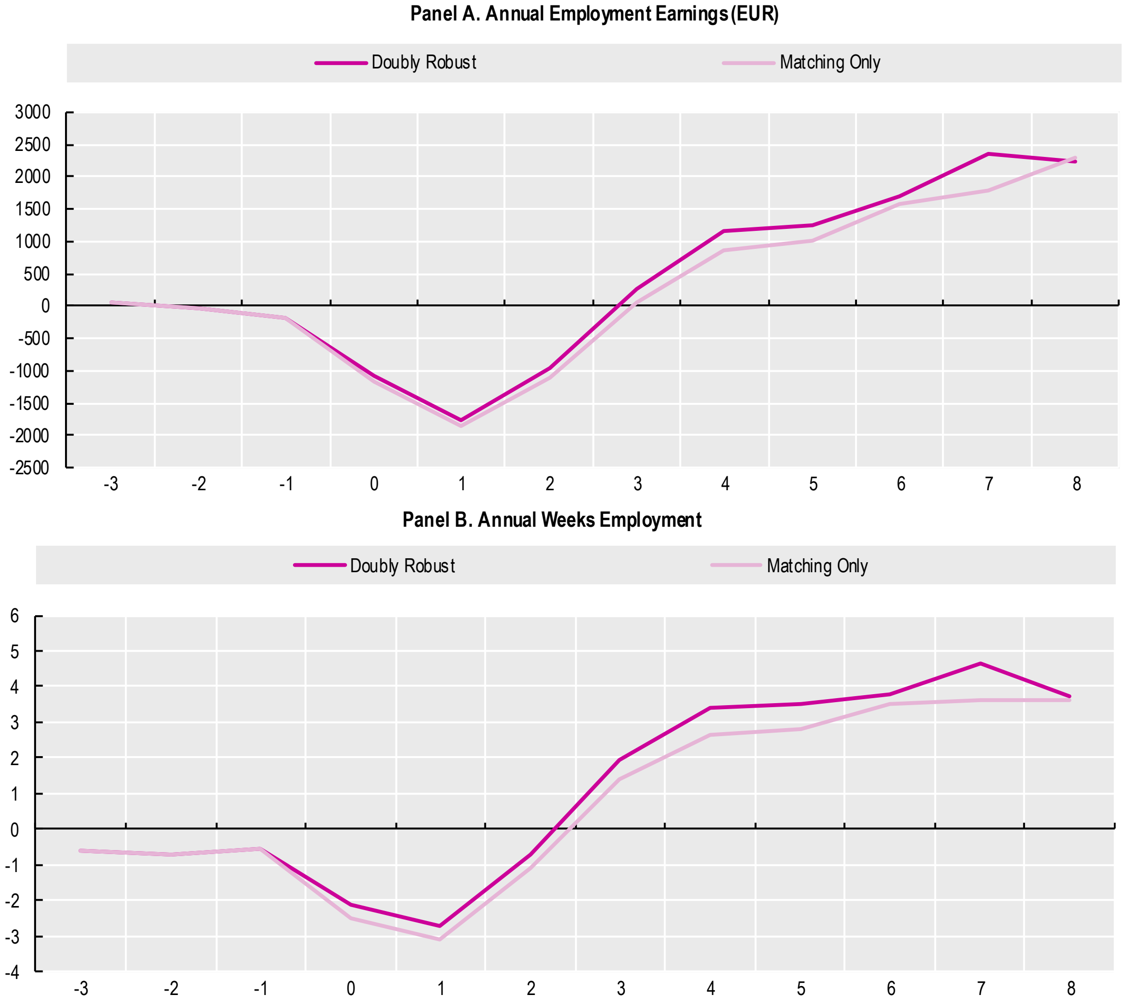
This means that where matching is unable to perfectly balance characteristics of CE participants and non-participants, utilising these same characteristics in regressions to estimate the impacts of CE allows any remaining differences to be accounted for. For example, suppose after matching, the group of CE participants still had less prior employment than the group of non-participants they were compared to. By using these employment histories in the regression to estimate the effect of CE on future earnings, then the extent to which prior employment influences future earnings is accounted for, and final "balancing" of this characteristic is achieved via the inclusion of this variable in the earnings regression. This doubly robust technique is implemented for both the baseline nearest-neighbour matching and for the kernel matching sensitivity analysis, therefore it mutes some of the potential room for bias that kernel matching has introduced as it is not able to significantly bring down the overall standardised mean bias of the characteristics of CE participants and their comparison group.

Doubly robust adjustment shows that matching has baseline performed well

The headline results in the main report employ a second adjustment after matching, known as a "doubly robust" adjustment. This involves estimating a regression on the matched sample to incorporate individual characteristics that may impact upon future outcomes. To the extent that the matching used to evaluate similar participants and non-participants is estimated using the same characteristics, if this matching has been successful, there should be little impact of employing the doubly robust adjustment. The advangtage of using the doubly robust adjustment is that where matching has not been able to completely remove all differences in the characteristics between individuals, the second stage adjustment can attempt to remove the residual differences. It is for this reason it is employed as the baseline technique in the main report.

Figure 4.1 confirms that incorporating the doubly robust adjustment makes only modest impacts on the overall results. The results using matching alone, or using matching and a doubly robust adjustment (the "baseline" technique used for analysis) are very close to one another. For brevity, results are reported for annual employment earnings and weeks of employment only, but across all other outcome variables in the report there is a similar picture. The one exception to this is in the results in year 8 for disability benefit. The impact for matching alone is -6.8%, whilst for the doubly robust specification it is -8.6%. However, the particular sensitivity for the disability benefit impacts is discussed elsewhere in this report and the headline impacts of the two different specifications provide the same underlying narrative. Differences in previous years are smaller in magnitude.

Figure 4.1. Performing a second "doubly-robust" regression adjustment makes little impact on results

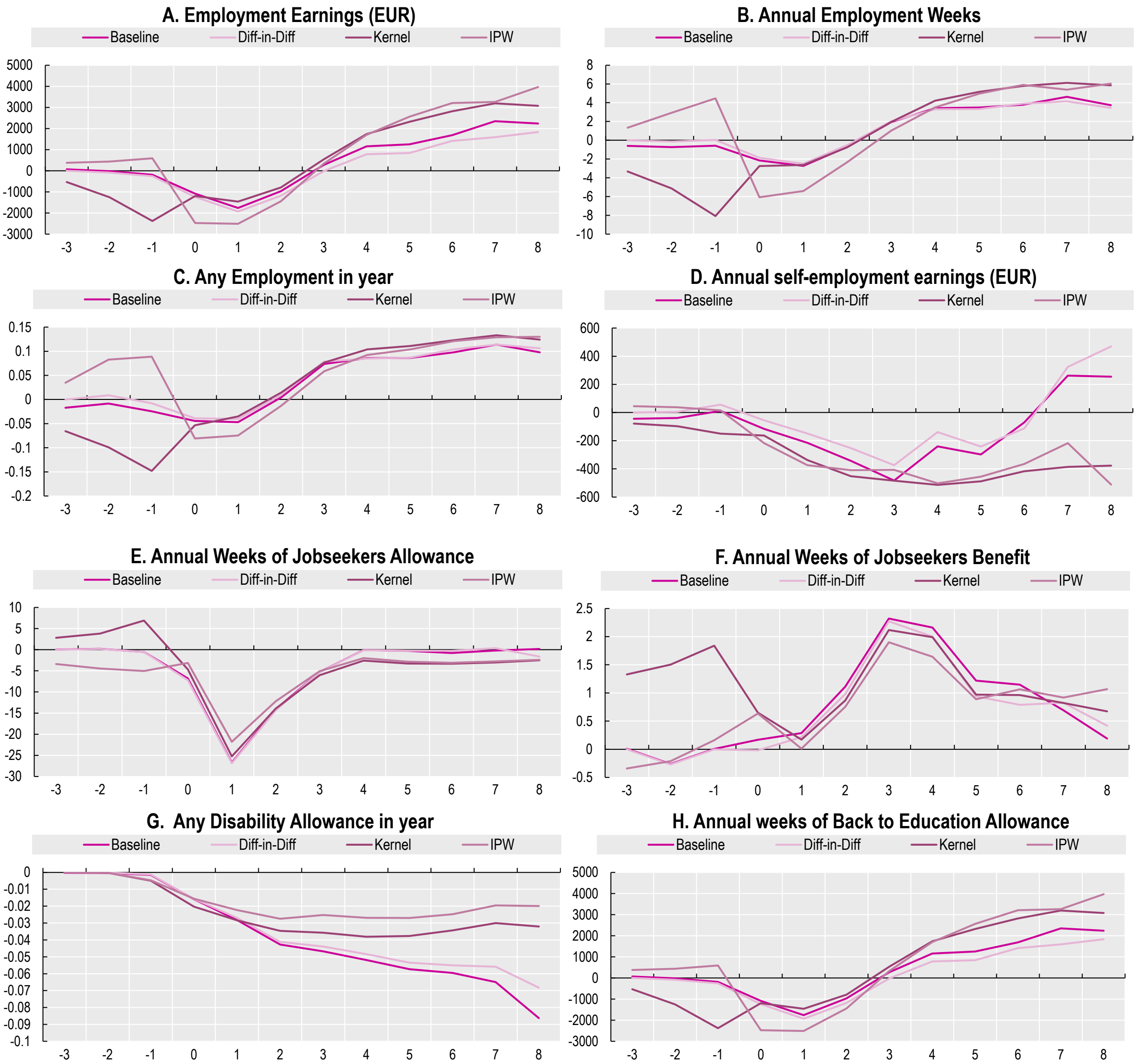


Source: Analysis of Department of Social Protection and Revenue Ireland data

4.1.2. Results tell a similar headline story across different analytical specifications

When reviewing the impact of different specifications on each of the main outcome's variables, the high-level narrative that the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) outlines is unchanged. But it is important to contextualise this somewhat in the variance of the different outcome estimates of the sensitivity analyses. This concordance in overall narrative is not to say that the estimates do not vary. For example, in Figure 4.2 panel A shows that in both the Kernel and IPW specifications, longer-run impacts on annual employment earnings are around EUR 3 000, this is higher than the EUR 1 500-2 000 level suggested by the baseline specification and the DID variation. The impact during the "lock-in" period (years 0-2) is similar across the specifications, qualitatively larger only for the IPW specification (at around EUR -2 500, compared to EUR -1 500 for the others).

Figure 4.2. Results tell a similar story across different identification strategies.



Note: Baseline refers to 1st nearest-neighbour matching with replacement. Diff-in-diff combines baseline matching with outcomes differenced relative to year minus 3. Kernel matching using Epanechnikov kernel and a bandwidth of 0.06. Common support is applied to all quarterly matched cohorts and in IPW re-weighting.
Source: Calculations based on Department of Social Protection (DSP) data.

However, it is important to remember the overall uncertainty in these estimates. For the baseline specification, the confidence interval averages around 1 100 EUR in years 6-8 for annual employment earnings, and grows over time as the sample size reduces. Given that this report also finds some variation when looking at annual cohorts (see section 4.2.1), and that the composition of today's CE claimants, today's CE schemes and today's labour market, will be not be exactly the same as those analysed in this report, it is more helpful to think of the results in their broader and more strategic sense. In terms of employment earnings, this higher-level picture is that earnings are reduced in the shorter term as participants take part in CE, before rising above those of their similar non-participant cohort over the longer term. To the extent that cumulative earnings are higher overall for CE participants by the end of the analysis period. Whether the annual impact is precisely 2 000, 1 500, or 3 000 EUR is most likely a degree of certainty too far. But what these variations in technique demonstrate is that the strategic overview given by the headline results holds true regardless.

The impact on annual employment weeks (Figure 4.2, Panel B), tells a similar story to that on annual employment earnings. Both the Kernel and IPW estimates show a higher point estimate (around six extra weeks of annual employment) in the longer term, than the baseline and DID specification (around four extra weeks- though there is less difference between these two techniques for this outcome, relative to their difference on annual employment hours). This interpretation persists when looking at any employment in the year (Figure 4.2, Panel C). Again, both the IPW and Kernel specifications report slightly larger impacts (around 12.5 percentage points increase in years 6-8 against 10.5 percentage points for the baseline and DID).

When looking at the receipt of unemployment-related benefits, Jobseeker's Allowance (JA) and Jobseeker's Benefit (JB), again the results suggest a concordance between the techniques, although with some slight nuance. For JA the long-term impact for both the baseline and DID specification is centred around 0 and is statistically insignificant. However, both the IPW and Kernel specifications see reductions to the annual receipt of JA (averaging 2.7 fewer annual weeks of receipt across years 4-8). Given the pre-period trends of both the baseline and DID specifications are better balanced, it is tempting to give these greater weight in the assessment of longer-term impacts, but given the concomitant improve to employment outcomes, there is room for cautious optimism of the possibility at least of slightly improved outcomes for this variable over the longer term. Differences in the early outcome period (years 0-4) are less apparent between the specifications. For JB, broadly all of the specifications paint a similar story, albeit with slightly higher variation due to the fewer individuals with any JB receipt in a given year.

4.1.3. Disability allowance receipt and self-employment do exhibit some variation

Two outcomes where it is worth pausing for a moment and considering potential differences in estimates across specifications are disability allowance receipt and self-employment earnings.

On disability allowance for example, the point estimate impact in year 7 ranges from -6.5 percentage points (baseline specification) to -1.96 percentage points (IPW). On its own, it might be reasonable to conclude that, similar to the considerations above, the baseline and DID specifications should be given precedence. However, for this particular outcome, all specifications coalesce around 0 in the pre-treatment (years -3 to -1) period, so there is no clear "winner" in this respect (they all produce good balance in the pre-eligibility time period). Furthermore, the results presented later, on annual cohorts (section 4.2.1), show a similar degree of variation in outcomes at longer-term periods. So it is prudent to be more circumspect on the precise impact of CE on longer-term impacts on disability allowance receipt. It is again important to remember the overall uncertainty in estimates at this point (the confidence interval on the baseline specification goes from -4.55 to -8.45 percentage points). But that broader thread on "high-level" narrative remains- whichever analytical technique might best represent reality, they all provide estimates that run in the same direction- that CE participation leads to lower levels of disability allowance receipt.

The result is not so clear cut on the impact of CE on self-employment earnings. The Kernel and IPW specifications suggest that CE participation negatively affects self-employment earnings, at levels that are statistically significant from year 0 in the time horizon onwards, never recovering from the "lock-in" period of CE. This contrasts with the baseline and DID specification which see a gradual recovery from their nadir in year 3, so that they turn positive by year 7 (and are statistically significantly positive in this year). There are relatively few individuals with positive self-employment earnings per year, which means of the variance estimates is high (the 95% confidence interval for the baseline specification is EUR -143 to ERU 653 in year 8). The small number of individuals affected, compared to, for example, the impact on employment earnings, means that such uncertainty does not have as large an impact on our overall narrative of CE. But it does mean that some reflection must be given as to whether the evidence generated in this report is sufficient to be overly assertive or prescriptive when it comes to self-employment.

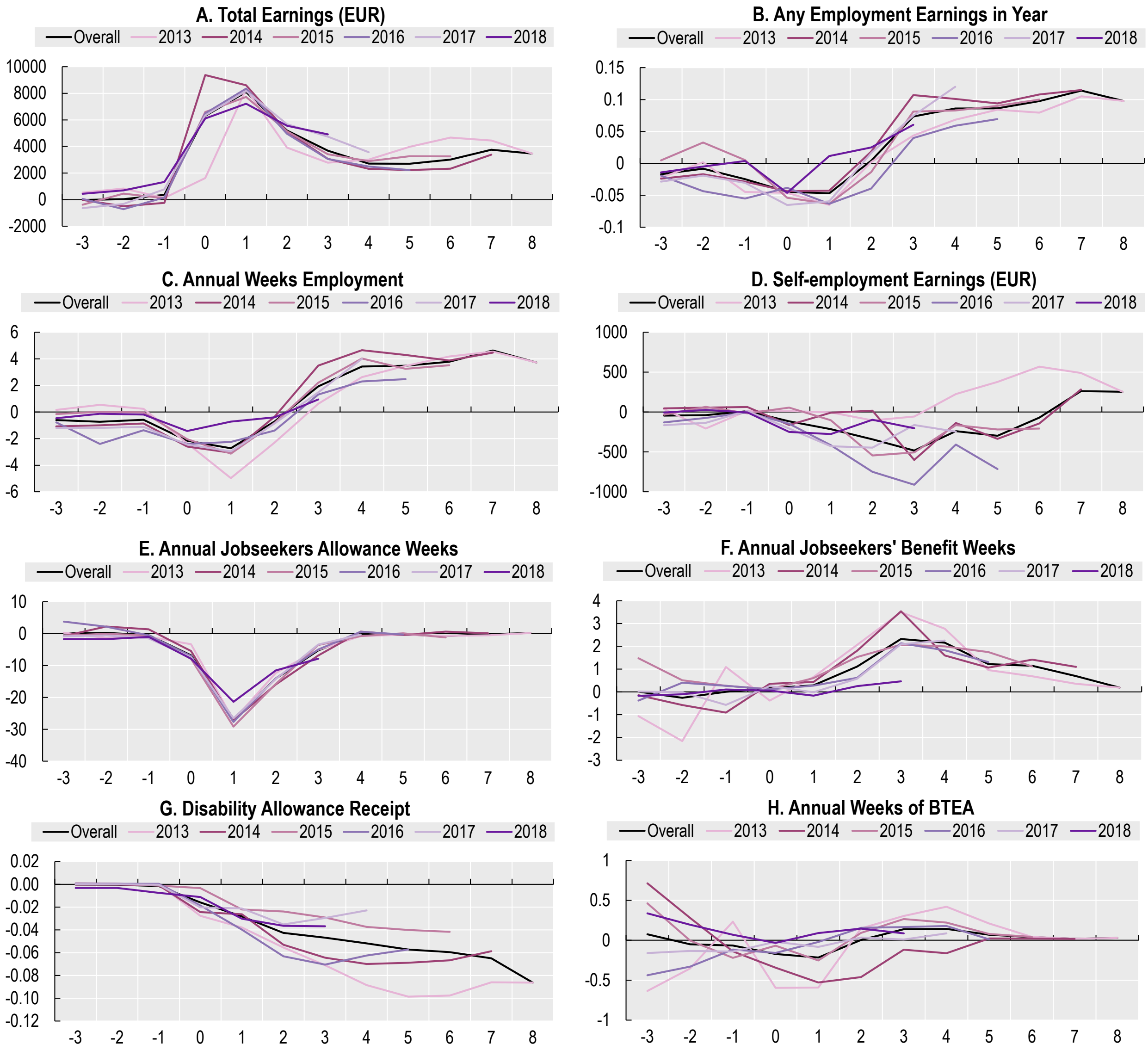
4.2. Utilising more data or combining data differently does not impact headlines

Both of the sensitivity analyses conducted on different data samples suggest that the headline story that is derived from the baseline specification is robust to changes in the data utilised, across the different measured outcomes. This is confirmed both when reviewing the annual cohorts of CE participants separately, compared to when they are pooled together in the baseline specification, and by looking at a sub-group of individuals for whom a detailed risk-scoring questionnaire (PEX questions) is available, which provide further information on socio-economic characteristics and motivations for job-search activity.

4.2.1. Grouping analysis by individual years suggests central analysis is well specified

As Figure 4.3 demonstrates, across all of the different outcome variables used in the report, looking at individual annual cohorts does not produce markedly different impact estimates compared to the baseline specification which groups annual cohorts together. This has two implications. Firstly, it suggests that the baseline results used in the main report produce a coherent and consistent story of how CE helps individuals to engage with the labour market. Secondly, it gives some reassurance that utilising cohorts from (at most) ten years prior to now can still provide meaningful insight into how CE participants may benefit from CE participation today. However, this is not without some caveats, that will now be discussed below.

Figure 4.3. Effects for individuals year follows similar patterns across most outcome variables



Note: Overall relates to the pooled baseline analysis utilised in the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), which estimates a pooled model grouping all of the years together
Source: Calculations based on Department of Social Protection (DSP) data.

The similarity between the individual annual cohorts and the pooled baseline analysis suggests there is not a large amount of mis-specification in the baseline specification related to annual variations. Generally all of the estimates for the outcomes in Figure 4.3 show that the pooled baseline analysis (the “overall” series) is roughly situated in the middle of the individual annual cohort estimates. This provides re-assurance on the specification of the pooled model. The pooled model estimates participation probabilities on CE for all individuals across all years and does allow for the contribution of different attributes to differ across years.

For example, in the pooled model, if a female had a higher probability of CE participation, that probability would be similarly higher across all different cohort years. By contrast, the estimates for each of the individual annual cohorts utilise the same model specification but are estimated only for that specific annual cohort. In this manner, the effect of being female on CE participation could be different in 2013 to the effect of being female in 2018. To the extent that the effect of different characteristics did vary across years and this variation caused material differences in estimated participation probabilities, this could mean that individual annual cohort effects look very different to the pooled baseline analysis. This would be due to mis-specification of the pooled model, as annual variation in across explanatory variables was not reflected in the overall estimation.

However, this mis-specification does not appear to be present in the CE analysis conducted and there is a good degree of concordance between overall pooled baseline analysis and the individual years. This offers both reassurance in terms of the confidence in the overall results, but it also useful in terms of the analysis' "generalisability". That is, the more general specification of the baseline model- which is not unique on specific years- is preferable in the sense that it is more likely to provide a valuable benchmark for which to evaluate future CE cohorts. It is by no means certain and there can always be reasons why the future does not look like the past, but this more general specification provides as much guard against this as is possible. It provides this reassurance without a large cost to accuracy potentially lost through more disaggregated annual model specifications.

Although the narrative of the analysis remains unchanged, there is some variation in the results by year of the different annual cohorts. For example, for any employment earnings in the year, Figure 4.3, Panel B shows that in year 3 the individual annual cohort point estimates range from 3.96% for the 2016 cohort, to 10.7% for the 2014 cohort. This range is wider than the confidence interval for the baseline pooled analysis (a point estimate of 7.4% with a confidence interval of 5.9% to 8.8%). The interpretation of this is then that the baseline pooled specification presents a good "average" guide to expected impacts of CE. If the question was "what do we expect an average CE participant to experience in terms of labour market attachment three years after their first scheme, at an unspecified point in the business cycle" the pooled baseline specification will provide a good answer. However, when looking a specific individual, at a specific point in the economic cycle there may be particular factors that mean their impacts are greater or lesser than that average estimate. Therefore, interpretation should be slightly more circumspect when attributing effects on more specific cases (for example saying "individual X in year Y will earn this much extra"). However, the individual annual cohort estimates can still provide extra insight on the degree of uncertainty expected, which is not well captured by the confidence intervals on the overall baseline specification. As is to be expected, this variation is larger for those outcomes whose impact is determined by fewer individuals in receipt (for example, see self-employment, jobseekers benefit and BTEA – Figure 4.3, Panels D, F and H).

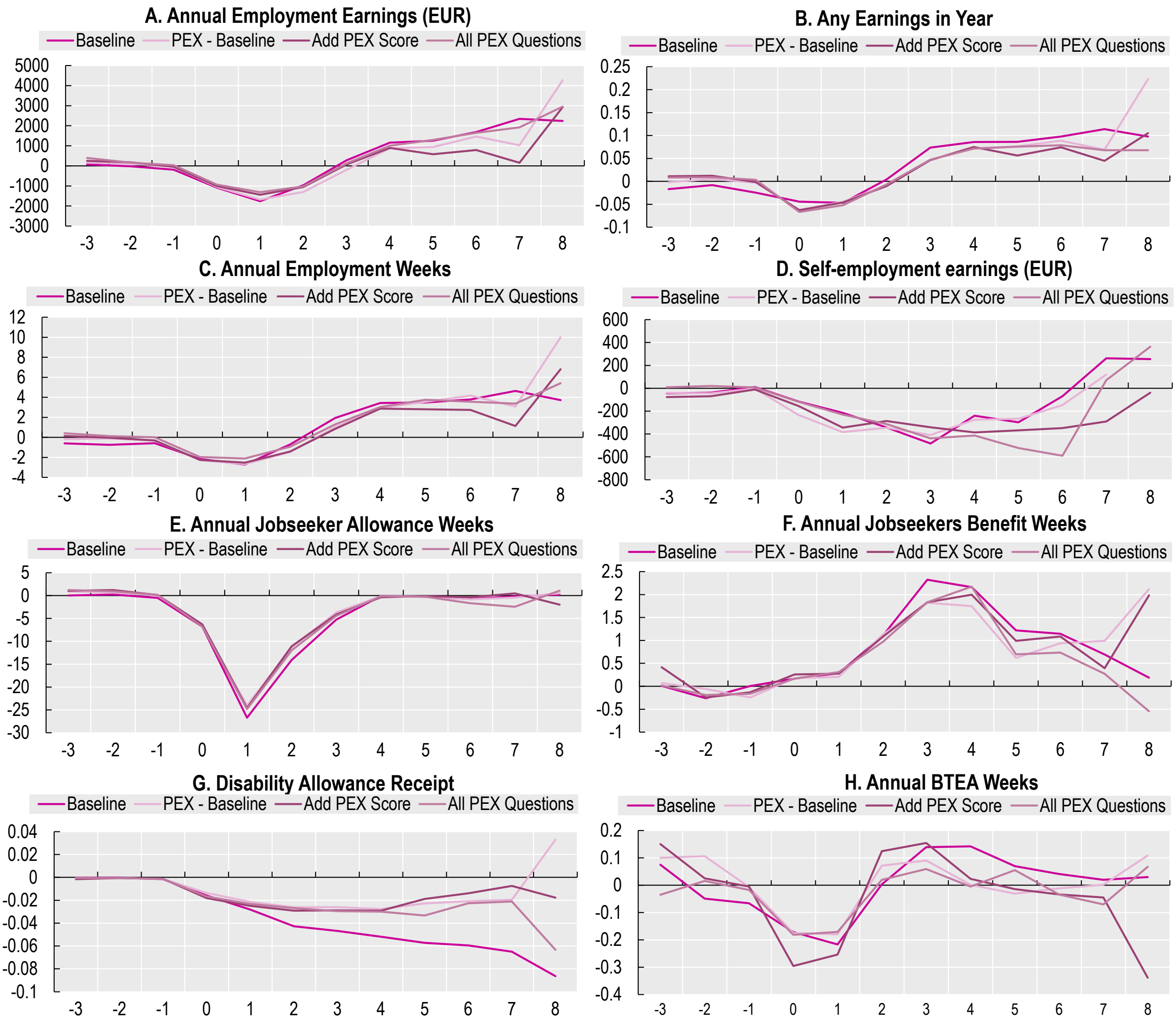
The annual variation for disability allowance receipt is starker than for other outcomes

Whilst the overall specification for disability allowance does a good job in averaging individual annual cohorts, there is a greater underlying variation of these annual cohorts. Figure 4.3, Panel G shows that in year 4, despite an overall estimate -5.7 percentage points the estimates range from -4 percentage points for the 2015 cohort to -9.9 percentage points for the 2013 cohort. A visual inspection of the different panels in Figure 4.3 shows that this variation tends to be consistently larger than other outcomes across the estimation horizon (self-employment earnings in Panel B with perhaps the most similar variation). This variation leads to the conclusion that the results are more sensitive to the precise sample used and that caution should be employed about interpreting exact magnitudes of estimates. However, the pattern of impact (all annual cohort display reduced receipt of disability allowance) provide reassurance on the overall narrative that CE leads to reduced disability allowance receipt.

4.2.2. Additional PEX data permit more complete investigation into omitted variable bias

The ability to analyse a subset of individuals for whom information is available on a suite of attitudinal and socio-economic questions related to potential labour market outcomes means more consideration can be given to how representative the baseline specification might be of actual programme impacts. The charts in Figure 4.4 present four combinations of model specification and data.

Figure 4.4. Adding PEX information demonstrates model stability



Note:
Source: Calculations based on Department of Social Protection (DSP) data.

The “baseline” specification presents the headline results from the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]). The “PEX – baseline” series then represent this exact same model specification, conducted only for those for whom PEX questionnaire information is available. The extent to which the series change between the baseline specification and the PEX baseline represents the extent to which the sample of individuals change, as the model to select individuals remains unchanged. PEX information is not randomly available. In particular there is better coverage of these data for later annual cohorts and there is a bias towards individuals with shorter unemployment claims (because those individuals that have continued long-term unemployment claims do not leave the Live Register. This means they do not have the “opportunity” to become unemployed again, at which point they would complete a PEX questionnaire). The “Add PEX Score” scenario works on the same PEX sample, but in the model specification adds one variable, the PEX score.

This is a numeric score given to individuals which summarises all of the information in individuals' questionnaire responses into a single index. This index relates to an individual's risk of still being unemployed after one year. In this sense, for the sample in this report, it is somewhat meaningless- the sample is, by definition, already unemployed for one year. However, this aggregate score may still convey information on the likelihood of CE participation and future outcomes, depending on the score's correlation with these factors. The final specification "All PEX questions" relaxes this relationship between the PEX score and correlation with CE participation and instead asks the question of how each of the individual questions may be related solely to subsequent CE participation probability. This is independent of any notion of their ability to separately predict the chance of an individual becoming long-term unemployed in the first instance.

The ability to utilise these extra PEX variables comes at the expense of sample size, particularly in later years. Table 4.1 demonstrates that by year 8, performing analysis on the PEX sample reduces the number of sample cases to 600, of which half are CE participants. This means that the confidence intervals for the PEX estimates are much wider than in the baseline analysis. For example, in year 8, the confidence interval for employment earnings in the PEX sample is 2.5 times as large as the confidence interval in the full sample of the baseline analysis. This should be borne in mind when reviewing the information contained in Figure 4.4. It is evident that in year 8 in many of the series there is an abrupt change in the series, which is more likely related to model instability and small sample sizes than it is to actual programme effects. For example, in Panel G on disability allowance receipt the two specifications for "PEX baseline" and "All PEX questions", which are almost identical in year 7, diverge suddenly in year 8 (PEX baseline moving from – 2.0 percentage points to 3.3 percentage points and All PEX questions moving from -2.1 percentage points to -6.3 percentage points). In fact, a graphical inspection of Figure 4.4 suggests that for sample of the outcome variables studied, this divergence starts occurring around year 7.

Table 4.1. PEX information comes at a cost of sample size

Year (relative to CE eligibility)	-3	-2	-1	0	1	2	3	4	5	6	7	8
Baseline	35 600	35 600	35 600	35 600	35 600	35 600	35 600	29 700	23 500	17 300	10 900	4 700
PEX	22 800	22 800	22 800	22 800	22 800	22 800	22 800	18 300	13 500	9 000	4 400	600
PEX percentage of baseline	64%	64%	64%	64%	64%	64%	64%	62%	57%	52%	40%	13%

Note:
Source: Calculations based on Department of Social Protection (DSP) data.

The usefulness of this extended set of PEX questionnaires, is their potential to offer insight into variables that are usually unobserved in more traditional administrative data that simply detail receipt and histories of government benefits. For example, there are questions on the willingness to move for work, self-assessed difficulty with reading, writing and numbers, on the ability to speak English, on access to use of own or public transport. These questions provide the ability to have a much greater assessment of an individual's motivation to find work, on their ability to undertake work and on their access to jobs. They provide a much richer picture of an individual that is possible to derive from those traditional administrative data. To the extent that these variables do influence both the decision to participate in CE and subsequent labour market outcomes, the baseline data sample accounts for them only through the ability of the existing variables to proxy them. For example, if greater motivation to seek work is also associated with better previous earnings, then only having access to data on previous earnings will allow some of the influence of that motivation to be controlled for via the incorporation of those "observable" characteristics in the administrative data that the estimation can make use of.

Additional PEX data doesn't make significant differences compared to the baseline specification

When looking at whether or not the additional PEX information provides extra explanatory power to match individuals on information that was previously unknown, the results are similar to the annual cohort analysis. In Figure 4.4 the difference to consider is not the gap between the "Baseline" specification and the PEX specifications, which represent changes in the sample, but to compare the "PEX Baseline" with "PEX Score" and "All PEX questions". If these series are similar (particularly the "PEX Baseline" compared to "All PEX Questions"), then it means that the addition of all of the extra contextual information from the PEX questions has had little impact on the overall results. For the most part, this extra information does not greatly change the headline results and suggest there is limited omitted variable bias in the baseline specification (which evaluating what extra explanatory power that the PEX questions may bring). This is a similar finding to the broader academic literature reviewed in section 3.1.2.

For example, six years after treatment, the baseline specification on the PEX sample provides an estimate of the reduction in disability allowance of 2.1 percentage points, adding the PEX score reduces the impact to 1.4 percentage points, whilst adding all of the PEX questions increases it back to 2.3 percentage points. These results are similar to one another (their confidence intervals overlap each other) and compare to the impact from the baseline sample in the main report of 6 percentage points. Similar results manifest themselves for the other outcomes studied in Figure 4.2.

The PEX disability allowance sample varies markedly from the baseline analysis

Despite making the argument that it is preferable to compare different PEX sample specifications, it is insightful to consider how disability allowance receipt varies in the PEX sample compared to the baseline sample. Changing to the sample of individuals with PEX information, using the original nearest neighbour matching specification for the main analysis results a much lower impact of participation in CE on disability allowance receipt. The effect reduces to roughly a 2-percentage point reduction, compared to around 6 percentage points in the baseline sample. This may be explained by the different dynamic of the sample of individuals with PEX information. PEX participants are more likely to be from recent cohorts and are less likely to be individuals with long durations of unemployment (because once an individual leaves unemployment benefits, if they re-appear on the Live Register they will have to complete the PEX questionnaire). However, it is important to review this evidence alongside that presented in 4.2.1 that showed variation in disability allowance outcomes by individual annual cohorts. The variation demonstrated there is concordant with the further variation here in moving from the baseline sample to the PEX sample. Overall this presents the picture that the results on disability allowance are quite sensitive to the precise sample used for the measurement, so caution should be employed over too tightly prescribing specific magnitudes of impact on disability allowance receipt.

Box 4.1. PEX data show promise for further sub-group analysis

In addition to allowing robustness checks on the data, the PEX questions allow the report to look at particular sub-groups of individuals that it is not possible to review in the main specification. For example, this sub-group analysis can review how well CE serves low educated individuals, or those without access to their own transport. This may have important policymaking benefits. If caseworkers know that particular individuals are able to benefit well from CE placements, they may be able to offer encouragement to those individuals to participate, in the knowledge that the scheme is well-suited to providing for them.

Better educated individuals do better on CE

It is difficult to determine precise impact for the PEX subgroups, owing to the very small sample available. However, one group stands out particularly among all of the subgroups as doing well after participation and that is those individuals with higher levels of education (Figure 4.5). In terms of employment earnings, the point estimate for this group is around twice as large as for the full PEX baseline sample, on any employment in the year the long-term impact is a positive 14 percentage points (against 7-8 percentage points in the PEX baseline), and on annual weeks of employment around 7 extra weeks compared to the PEX baseline of 4 weeks. These are large differences and suggests that where individuals with higher levels of education have experienced long-term unemployment, the beneficial effects of CE can unlock their latent labour market potential.

Less clear-cut sub-groups, though worth pausing to reflect on, are those with lower education and difficulty with reading and writing. The point estimates here are lower than for the overall PEX sample and whilst they may not be significantly different in a statistical sense, they may be worthy of further attention in a practical sense. Further evidence generation on the difficulties faced by these groups, and how CE might help them to achieve their goals in the labour market could help to focus the support that CE offers them and help to guide how the training budget may be better used to serve their needs. (OECD, 2021[27]) offers advice on how targeted and comprehensive support is necessary by PES to support these individuals back into labour market.

As PEX data coverage increases, further review this sub-group analysis

The limited sample sizes available for these groups should improve over time as more and more individuals have PEX questionnaires completed. Therefore, re-analysing groups in the future will be able to estimate more precise effects and further aid policy delivery and development.

4.2.3. Taken together the results show stability but some caution should be given to the scale of the impacts on disability receipt

The robustness checks confirm the primary conclusions of the main report, demonstrating that CE's effectiveness to improve employment outcomes and reduce disability allowance receipt remains robust across various methodological and data variations. These checks provide additional assurance that the observed impacts of the CE program are not merely the reflection of specific analytical choices but mirror the program's actual effects. However, when interpreting these results, particularly the scale of impacts on specific outcomes like disability allowance receipt, some caution is warranted. There is some sensitivity of the results to different sample compositions, highlighting the need for careful consideration when generalising the findings.

Annex – Further PEX samples

Figure 4.5. PEX subgroups – higher educated do better on CE

Source: Calculations based on Department of Social Protection (DSP) and Revenue Ireland data.
:

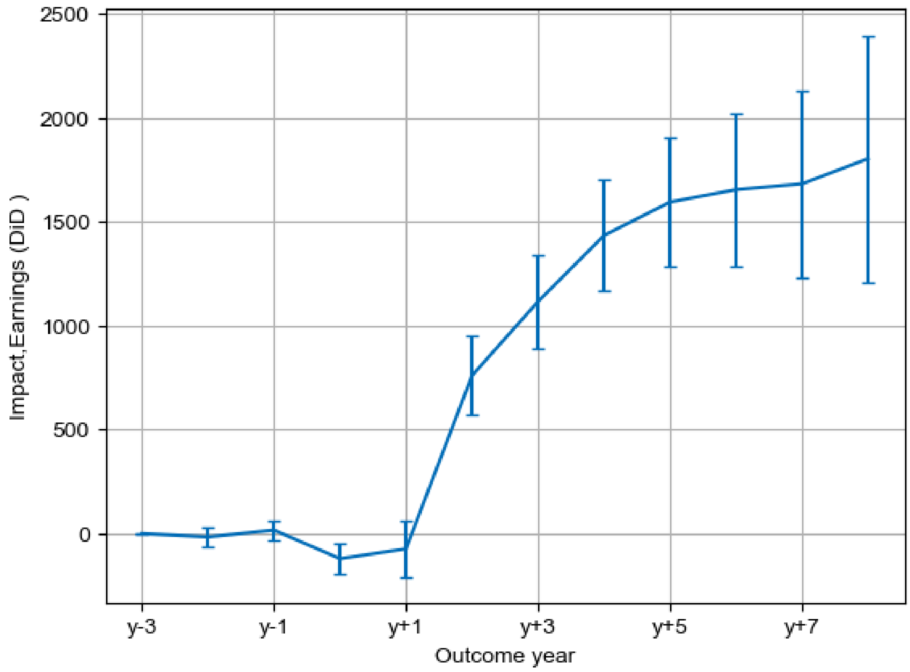
5 Tús Analysis Robustness Checks

As with the CE robustness checks on the main results, this chapter considers two aspects of the analysis of Tús. The first centres on varying the technique underpinning the baseline analysis and making more explicit the checks performed at each stage of the analysis. The second alters the sample and variables used to analyse the degree of variability in the result, compared to the baseline analysis used in the main report.

As part of the sensitivity analysis on the estimation technique used, difference-in-difference (DiD) technique is also applied to the Tús analysis. The weights generated in the initial phase described above are used again. Changes in the outcome variable (earnings) are compared relative to the value three years before participation or eligibility. This method checks for any pre-existing trends in the years prior to eligibility that could potentially lead to the difference in outcomes between participants and eligible non-participants. The rationale is that, even if the treatment and control groups differ in respect of the outcome variable, that difference should be similar in the absence of the treatment. Estimating the difference over time between the post- and the pre-period of the treatment group's outcomes, while controlling for the difference over time in the control group's outcomes, provides the casual impact of the intervention. In other words, even if there is some underlying difference between the two groups, this approach calculates only the difference over time relative to pre-existing difference. There is little variation in the impacts between the DiD approach and the baseline specification. This provides re-assurance that the controlling variables used in the baseline specification were successful at removing differences between the treatment and control groups.

The additional reassurance provided by the difference-in-differences approach is that, in its comparison of outcomes, it controls for possible pre-existing differences between the two groups in the pre-treatment period that do not change over time. This addresses any potential concern that there are time-invariant unobservable characteristics that systemically differentiate individuals in the treatment and control groups and that these are correlated with outcomes. Given the result, there is no reason to believe the main results are driven by such differences.

Figure 5.1. Impact of treatment on earnings (difference in differences)



Source: Calculations based on Department of Social Protection (DSP) data.

Other identification strategies were also considered. The instrumental variable approach seeks to find a variable that affects the selection into treatment but is not correlated with outcome variables, nor with the outcome variable (earnings or employment). Exploring the data, it was not possible to find an instrument that was suitable to control for selection into Tús but that did was not also related to outcomes. Improving referral data coverage may provide more recourse to such an approach, if it is possible to better determine how different Intreo offices make referrals of their customers. Similarly, with improvements in referral data consistency, exploiting the random selection aspect of the process remains an avenue for future evaluations of Tús.

5.1. Omitted variable bias – additional PEX variables make only a minor difference to results

This section considers the additional variables available in the PEX dataset to assess the extent to which the inclusion of factors such as education level, health status and an individual's willingness to change location for work can affect the magnitude of the impact estimates. Adding in PEX information for the subsample for which it is available facilitates analysis of the potential for omitted variable bias in the headline results. If there are minimal differences across the three specifications, then it can be concluded that the PEX information has little extra explanatory power beyond what is captured in the baseline specification.

Sub-group analysis of just those who have PEX information reduces the sample size by approximately 50%. Due to the timing of the PEX roll-out in 2012 (in the context of the analysis period), the make-up of the PEX subsample is distinct from the overall sample. This is an important factor which must be considered when interpreting the results that follow.

5.1.1. The first exercise applies the weights generated from the modelling of the full sample and considers the PEX variables only at the regression stage

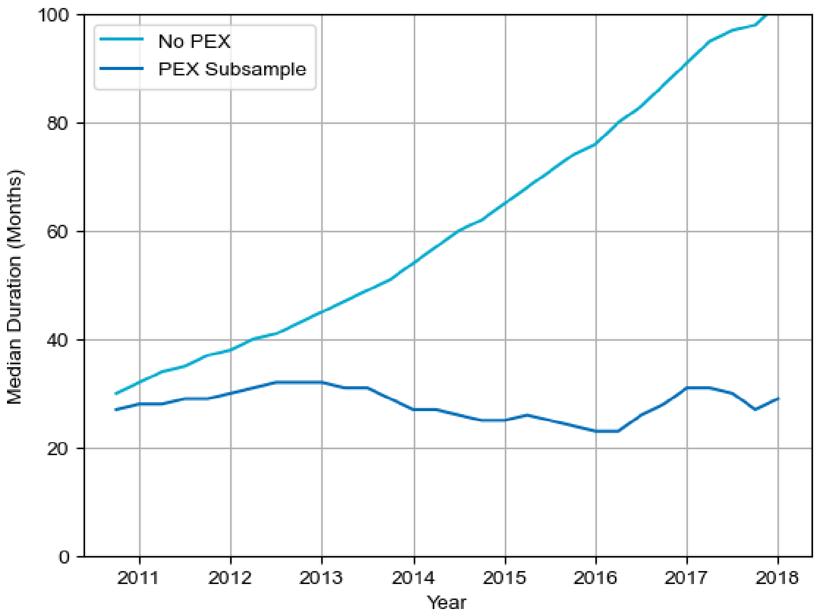
The “baseline” specification reproduces the headline results as seen in Chapter 7 of the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]). Recalling these baseline results, the effect of treatment on earnings from employment in the third year after participation is EUR 1 100. The effect increases to EUR 1 500 by the fifth year after participation. Using weights generated in the baseline model, the potential for omitted variable bias is assessed by comparing the “PEX Baseline” with “PEX Score” and “Add PEX questions”.

> *Outcomes for the subset for whom PEX information is available use the same model specification, in "PEX Baseline"*

The difference in outcomes between the headline results and "PEX Baseline" reflects the difference in the samples. Using the same model specification as applied to the complete sample, the effect of treatment on earnings is lower for the PEX cohort. In the third year after participation, the participants in the PEX subgroup experience an increase to employment earnings of EUR 886, in the 5th year after participation the effect grows to EUR 1 130 when compared to a group of eligible non-participants who also have PEX information.

That the difference in outcomes reflects underlying differences in the two subsets is worth reiterating. As illustrated in Figure 5.2 people without PEX scores have longer unemployment durations, particularly in the later quarters. Given that PEX was only introduced in 2012, the people without scores are more likely to be from earlier cohorts, where a claim was open before the practice of PEX scores began. Once an individual ends a claim and restarts a new claim, the PEX questionnaire is completed and a score generated. This implies that those without scores will have longer durations, as illustrated below.

Figure 5.2. Difference in median unemployment duration



Source: Calculations based on Department of Social Protection (DSP) data.

"Add PEX Score" models the outcomes for the same PEX sample, with the addition of one variable, the PEX score

The PEX score is a value between 1 and 99, which summarises the questionnaire responses into a measure of the probability of remaining unemployed one year after the claim start. The PEX baseline model is amended to include only the PEX score as an additional explanatory variable. When the PEX score is introduced to the model, the impact of treatment in the third year after participation is EUR 910 and in the fifth year is EUR 1 200.

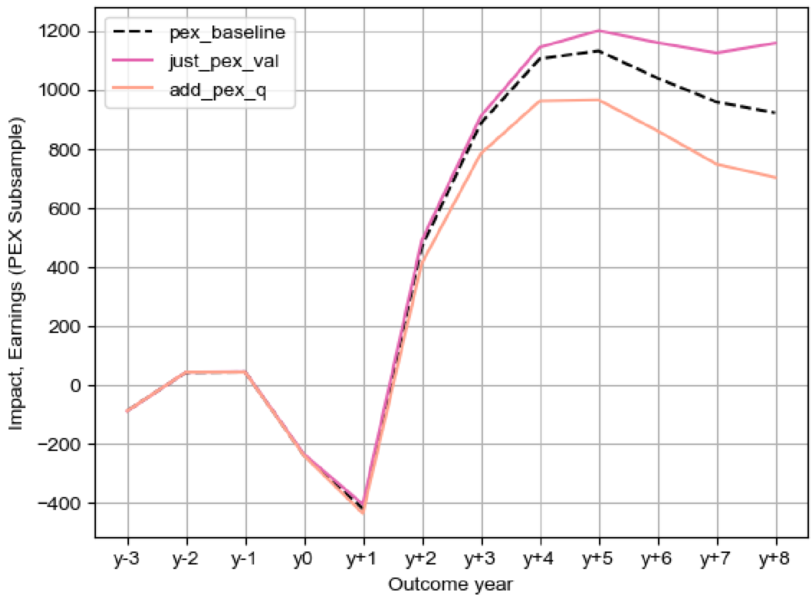
"Add PEX Questions" adds selected PEX variables to the outcome model for the PEX subsample

In addition to the PEX score, other variables included in this version of the participation model are: an individual's willingness to move for work (Yes/No), difficulty reading/writing (Yes/ No), Health Status (ranging from Very Good-Very Bad), Education level (ranging from lower-primary to third level).

Including additional PEX data does not make a significant difference to the estimation of the treatment effect on outcomes. In the third year after participation, there is a difference of EUR 24 between the "PEX baseline" and "add PEX score". Compared to the PEX baseline, adding PEX questions reduces the treatment effect by a similarly modest amount (EUR 103) in the third year after participation.

The larger differences between the baseline model results and the "add PEX score" or "add PEX Qs" results are not relevant in respect of omitted variable bias. As noted above, these reflect differences in the sample, which is illustrated across one dimension in Figure 5.3. Accordingly, the exercise presents little evidence that a relevant characteristic captured in the PEX variables has a substantial impact on the outcome models underpinning the main results.

Figure 5.3. Impact on earnings for PEX subsample



Source: Calculations based on Department of Social Protection (DSP) data.

5.1.2. Weights are estimated separately for the PEX subsample

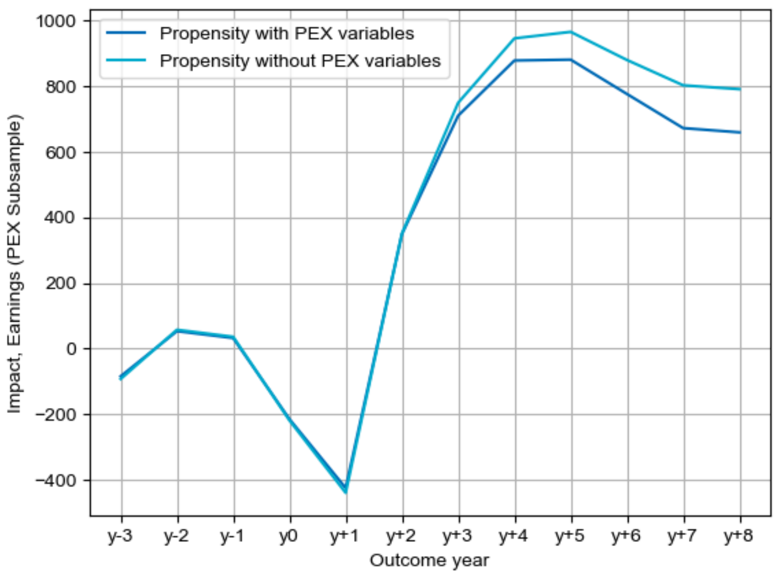
The second element for testing the possibility of omitted variable bias in modelling Tús participation treats the subsample with PEX information as entirely distinct from the overall population. The specification “Propensity with PEX variables” generates propensity weights with the model considering the additional PEX variables while the specification “Propensity without PEX variables” considers only the variables included in the baseline model outlined above. The PEX variables included are the same as those listed in 5.1.1.

After the weights have been assigned, the “Propensity with PEX variables” group includes the PEX variables as explanatory variables in the second stage Weighted Least Squares (WLS) regression. The “Propensity without PEX variables” group does not. The results are depicted in Figure 5.4.

When the PEX variables are not included for the PEX sub-group estimation, the effect of treatment on earnings from employment in the 3rd year after participation is EUR 749. The effect increases to EUR 965 by the 5th year after participation. When the PEX variables are included, the effect of treatment on earnings is reduced. In the 3rd year after participation, participants with PEX information experience employment earnings of EUR 710 higher than non-participants who also have PEX scores. The effect increases to EUR 880 by the 5th year after participation.

The impact on earnings for the sub-group with PEX information is considerably lower than the overall sample. This again reflects the association between the introduction of PEX in 2012 and the analysis time-period beginning in the second half of 2011 (and an eligibility requirement of one year of unemployment duration by that point). As evidenced in Figure 5.2 above, median unemployment duration for people without PEX scores is considerably longer than those with PEX scores, particularly in the later quarters. The people without PEX scores are more likely to have opened their claim before 2012 and those in the later quarters of the analysis period may well have been long-term unemployed for some time with limited labour market activity. These circumstances, rather than the existence of a PEX score, account for the variation in the treatment effect in the subsample.

Figure 5.4. Impact on earnings for PEX subsample (re-weighted)



Source: Calculations based on Department of Social Protection (DSP) data.

5.2. Results are robust to altering the sample by removing "future-treated" individuals

This sub-section looks at the impact of changing the sample of non-participants, to exclude those that will participate in Tús at some stage in the future.

5.2.1. The treatment effect on earnings decreases when "future-treated" individuals are removed from the control group

Where an intervention is delivered at a fixed point in time, individuals can be assigned as treated (participate in the programme) or not treated (do not participate). When enrolment is ongoing and the analysis period is long enough, identifying individuals as part of either the treatment or control group may be challenging - they may be first part of the control group and subsequently part of the treatment group. In this scenario, where any unemployed individual can potentially become a participant, a number of analytical approaches are taken to avoid conditioning on future states. For example, people who are not observed participating in the programme may remain untreated over the entire analysis period because they have found employment in the interim. Sianesi (2004) notes that choosing as the non-treated those who never participate in a programme amounts to conditioning on their future (successful) outcomes.

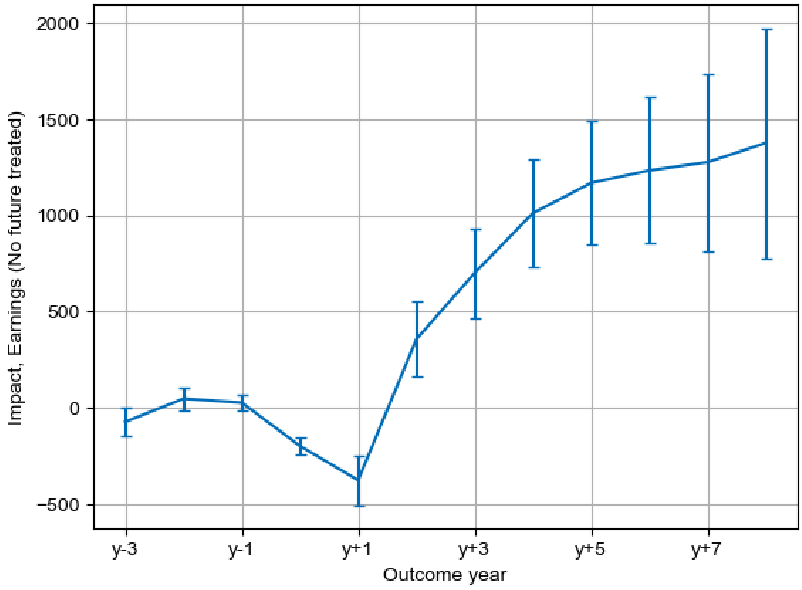
As noted in Box 6.1 of the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), the approach in both CE and Tús evaluations is to compare people who begin treatment with people who have been unemployed for a similar duration but have not yet entered the programme. This allows for the fact that, while they have not entered the programme in a given quarter, they may do so in a subsequent quarter. This is the central comparison that underpins the analysis, rather than a comparison of the treated with the never treated.

Being eligible for Tús requires an unemployment duration of at least 12 months. Given the long analysis period in this evaluation, the dataset includes people who may take up a Tús placement much later than the first quarter in which they become eligible (the greatest distance being someone who is long-term unemployed and eligible in 2011 and a Tús participant in 2021).

One option is to remove all such individuals from earlier control groups. For example, the evaluation of BTEA (Kelly, McGuiness and Walsh, 2015[28]) removes future treated individuals (control group individuals who subsequently undertook a BTEA course during the evaluation time period), control and treatment group individuals who went on to undertake a Community Employment (CE) or Back to Work (BTW) scheme during the evaluation time period, treatment group individuals who received unemployment training during the evaluation time period, and control group individuals who commenced a training course six months prior to the outcomes timeframe.

As a sensitivity analysis exercise, it is worthwhile considering whether the programme effects are entirely driven by this decision. By changing the sample to remove all individuals who are treated in future quarters, leaving just those who are never treated, the effect of treatment remains, as outlined in Figure 5.5. In the third year after beginning Tús, rather than former participants experiencing an increase of EUR 1 100 relative to non-participants, the earnings of the treated individuals are EUR 700 higher than non-participants. Similarly, in the 7th year following their Tús start date, former participants see their earnings rise to a level just over EUR 1 200 above their non-participant peers.

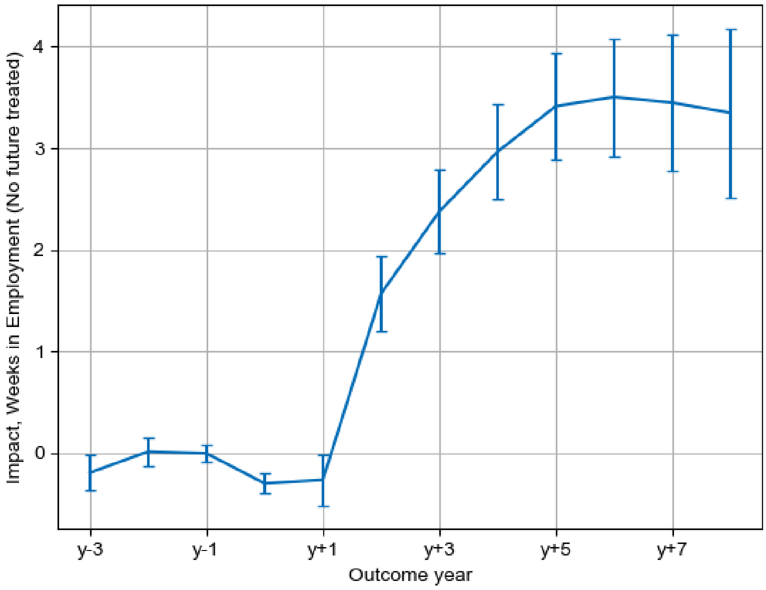
Figure 5.5. Removing no future-treated individuals lowers earnings impact but effects remain positive



Source: Calculations based on Department of Social Protection (DSP) data.

In the case of weeks worked, the results are similarly robust to changes in the sample (Figure 5.6).

Figure 5.6. Removing no future-treated individuals does not greatly impact weeks of employment



Source: Calculations based on Department of Social Protection (DSP) data.

5.2.2. The treatment effect on earnings decreases less when only "future-treated" individuals are removed who participate within 3 years

Arguably, the removal of everyone from the control group who will participate in Tús at some point in the future is an over-zealous measure in this case. In the example mentioned above of someone who is long-term unemployed and eligible in 2011 and a Tús participant in 2021, ten-year outcomes are not calculated in the analysis and so the removal is changing the sample to no particular methodological end.

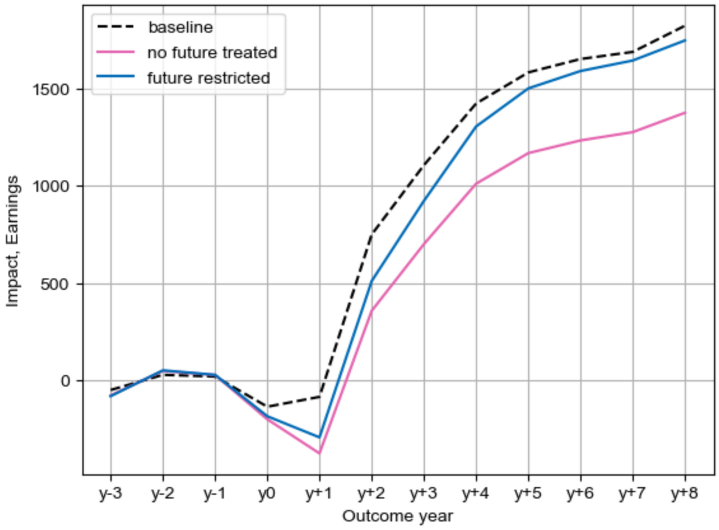
As an alternative measure, the restriction was amended to include in the control group in each quarter those whose Tús episode occurs at least three years from the quarter in which they are in the control group. Earnings information is available for all participants (members of the control group) up to three years after participation (eligibility). By restricting the inclusion of the future treated in each quarter to just those whose treatment occurs in at least three years' time, there is certainty that any observed differences in outcomes in that first three years cannot be attributed to any control group member's participation on Tús.

A further justification for this timeframe is that having completed a 52-week placement on Tús, a participant cannot subsequently re-participate on Tús for a minimum of 3 years (DSP, 2021[29]). By imposing the three-year restriction, the control group in this specification also excludes any individual who has just completed Tús, becomes eligible again through another spell of long-term unemployment before beginning another episode of Tús. This repetition of Tús over a short period of time with a brief window of being a control group candidate can make outcome results difficult to interpret, raising the question of what proportion of the impact on earnings can be attributable to time spent as a Tús participant and whether it was attributable to the first, second or both spells.

The three-year restriction offers a middle ground alternative for the removal of future-treated individuals in the quarterly comparison groups. The impact on earnings following this approach also falls in the middle of complete removal and no removal. Including those in the control group whose treatment occurs at least three years after that quarter results in a treatment effect in the 3rd year after participation of EUR 920 compared to EUR 1 100 when including the future treated and EUR 700 following their removal. The effect in the 5th year matches that of the baseline at EUR 1 500 and is higher than the effect realised when excluding the future treated of EUR 1 200.

While it is useful to vary the selection parameters for the comparison population for the purpose of sensitivity analysis, restricting the inclusion of control group members in this way fundamentally alters the sample analysed which is something that must be considered when interpreting the results above. The conclusion is that the results are robust to changes in the sample effected by the approach to those treated in future quarters.

Figure 5.7. Impact of treatment on earnings when removing future participants from the control group



Note: No future treated refers to removing all observations who are subsequently in the treatment group. Future restricted refers to removing from the control group in each quarter those who will be treated within three years. Baseline refers to including all of the future treated (as per headline results).
Source: Calculations based on Department of Social Protection (DSP) data.

6 Sequence Analysis Robustness Checks

6.1. Introduction

The present chapter provides the reader of Chapter 5 of the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) with technical details and robustness checks on the sequence analysis. More specifically, technical details will be related to data construction – to fully disclose the choices made to build sequences of employment and social protection provisions from the many raw data sources available.

In the analysis of sequences, data are rearranged so that every individual, at each month in the observed timeframe (i.e. from January 2012 to December 2019) are uniquely attributed to one among eight different "states": CE, Tús, employment with support (EWS), employment without support (EWoS), Back to Education Allowance (BTEA), JobBridge, JobPath, and a residual state capturing no participation in employment nor in any of the supporting schemes mentioned above. To this end, nine different sources of raw data – coming from 15 different underlying datasets – are merged and constructed into a grid where each row represents a person in a specific month of a specific year, while variables appear in columns. The next section will explain all the steps followed and methodological choices made in the creation of the final database.

Robustness checks, on the other hand, follow from data construction and methodological choices where some degree of arbitrariness was most needed. The spirit will be that of trying different – but still reasonable – assumptions, to see whether the results described in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) still hold. These arbitrariness turning points can be described as follows:

- In each month individuals in the data were assumed to belong to only one state, although states could be overlapping. To put it differently, states have been assumed to be mutually exclusive even if they might not always be. A recap of how often states overlap and which ones overlap the most will be provided.[1]
- Relatedly, not only mutually exclusive states have been assumed, but they also exclude each other according to a well-defined lexicographic ordering. A different ordering, also according to overlaps emerged under the former item, will be tried.
- The analysis in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) includes four different types of schemes under the broad state defined as employment with support (EWS). Among these schemes, casual jobseeker claims were included, i.e., unemployment claims under which jobseekers can work for up to three in seven days and receive an unemployment payment for the remaining days. The choice to include these claims under EWS was made to reduce the number of possible states, and therefore the complexity in the

analysis. For the sake of robustness, part of the analysis will be repeated assuming that casual claims represent a state on its own.

- As mentioned above, the data for the analysis in Chapter 5 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) was built with a monthly structure, therefore potentially allowing to analyse within-year transitions. However, apart from duration analysis, the sequence analysis in Chapter 5 of relies on annual observations. This choice was made to simplify interpretation while still retaining an overall view of the main patterns observed. In the present chapter a higher frequency – namely monthly – will be used, to show how much the choice of yearly intervals affected the analysis in terms of potential information loss.
- Two of the main data sources used in the analysis of sequences – namely, the earnings and contributions dataset, and the social welfare payments one – provide only aggregate annual records, with no information on start and end dates for benefit and employment spells. These datasets were the source of information for employment without support; most types of employment with support; as well as participation in other schemes, such as BTEA. The annual nature of these records implies that hypotheses were necessary to allocate employment spells and benefits within each calendar year consistently with data for which instead start and end dates are available. We describe the hypotheses made in Chapter 5 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) below, and further different assumptions about the allocation of employment weeks are put forward.

6.2. Data construction in depth

This section outlines how the data were brought together to identify discrete states for individuals at different points in time, so that sequences of different states could be observed.

6.2.1. Constructing eligibility spells for Community Employment

The first step in data construction is processing of the dataset holding eligibility spells of CE. The raw data count 749 010 episodes, including some spells which do not occur within the timeframe of the analysis in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]). After dropping duplicates in terms of individual identifiers, start and end dates of eligibility – and a few observations with missing start and end dates – 732 433 observations are left. After further merging consecutive spells with a gap shorter than thirty days, 731 203 episodes survive in the data. To allow some slack on eligibility end dates, each eligibility spell is extended by ninety days.

During this process, consideration is given to how Tús participation interacts with CE eligibility periods. CE eligibility periods are combined with Tús episodes. If a Tús episode starts within 30 days of the end of a CE eligibility spell, the original duration of the CE episode is extended until the beginning of the Tús episode, and the extra ninety days are added to the end date of the Tús episode. If a Tús episode starts between thirty and ninety days after the end of the CE eligibility spell, the extra ninety days are split before and after the Tús episode. These operations lead to the creation of some “new” short eligibility intervals, so that the total count grows to 760 637.[2]

Next steps concern the 488 054 episodes of CE participation that appear in the raw data. After i) dropping duplicates in terms of individual identifier, start date, end date, adult and child dependents; ii) merging episodes with intervals shorter than thirty days; iii) dropping episodes shorter than thirty days; iv) episodes held as supervisors, only 97 587 CE episodes survive. Out of these, 79 478 insist for at least one day within the target timeframe. Last, since CE is subject to a cap (see Chapter 3 in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), it is computed using a perpetual inventory method. Namely, as the period of interest starts in 2012, first the stock of accrued years under CE is computed until the end of 2011. This is performed in two steps. First, weeks of A8 or A9 contributions (which count annual payments of CE, Tús and the Rural Social Scheme, RSS) are used until 2010. Tús was introduced in 2011, so contributions to that point were mostly due to CE episodes, involving RSS a minority of individuals. Then, the stock of CE weeks accumulated until the end of 2010 is updated with weeks observed in CE episodes in 2011, to obtain the stock accrued at the end of 2011. This last stock is then updated month by month with CE episodes observed from 2012 onwards, to obtain a month-specific number of years spent in CE. Once this piece of information is combined with age, it is easy to identify the moment whence the cap is reached.

6.2.2. Constructing eligibility spells for Tús

Tús data are subject to a comparable cleansing procedure. The main difference is that instead of dropping supervisors, self-referrals to Tús are dropped. Also, a cap for Tús does not exist. Instead, a general restriction to Tús participation of three years exists after each Tús episode lasting 52 weeks. Tús episodes either stop immediately or la st until their (in most cases) maximum duration of one year. Consistently with CE, all episodes lasting less than 30 days are dropped. This means that a simple three-year restriction after each Tús episode survived in the data is introduced. Raw data include 52 257 Tús episodes, of which 49 379 survive in the relevant timeframe.

6.2.3. Information on other states is collated from different data sources

Information on other states is retrieved from four different data sources: the Department of Social Protection (DSP) benefits data, social welfare payments data, earnings data (from Revenue), and JobPath history data. These datasets can be broadly grouped into two categories, based on whether they contain records relative to spells – with start and end dates for each single episode – or annual records.

DSP benefits data and JobPath history data belong to the first category. Both contain detailed information on exact start and end dates of DSP benefits and JobPath referrals, making it easy to assign the relevant states to specific points in time during each year. The DSP benefits dataset is the main source of information for JobsPlus participation, i.e., for one of the schemes that falls under the supported employment status. It also provides information on individuals who, during the period of analysis, have been in receipt of maternity/paternity benefits, carers' allowance, disability allowance, pensions. JobPath history data show exact periods when individuals were referred to this employment service.

These two datasets only needed a limited amount of cleaning. For both, multiple episodes for the same individual were merged whenever the gap between the two was within 30 days. For JobsPlus, episodes with negative durations were dropped. For JobPath, missing end dates for episodes for which a starting date was registered, were replaced to have a one-year duration of the referral.

The earnings and contributions and the social welfare payments datasets posed a different set of challenges. The former is the source for information on employment without support. For the purposes of the analysis presented in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), the social welfare payments dataset represents the main source for data on receipt of Back to Education Allowance (BTEA), three of the supported employment types (Back to Work Enterprise Allowance – BTWEA, Back to Work Family Dividend – BTWFD, Part-Time Job Incentive – PTJI), and JobBridge.

As mentioned above, these two datasets are only available on an annual basis, with no information on start and end dates for benefit/employment spells. In order to reconcile this with the monthly structure of the dataset built for the analysis of sequences, some rules of thumb had to be implemented to allocate weeks of benefits/employment to specific weeks of the year. In particular:

1. When weeks of payment for supported employment, BTEA or JobBridge were observed in two (or more) consecutive years, a unique spell spanning over the two (or more) years was assumed. This

meant that weeks observed in the first year of the alleged spell were attributed to the last months of the year, while those observed in the last year of the spell, were allocated to the beginning of the year.

2. When weeks are observed in a single isolated year, a single spell was assumed, with a random start date (uniform distribution) so that the end date could be set by the end of the year.
3. If some employment without support was observed in the year when a period of supported employment was assumed to start (based on the rule under the first point above), its starting date was set at January 1st.
4. If some employment without support was observed in the year when a period of supported employment was assumed to end (based on the rule under the first point above), its ending date was set at December 31st.
5. In the other cases, supported employment was assumed to start on January 1st, and employment without support to end on December 31st.
6. Whenever the number of weeks associated to a payment or to employment was above 48, the whole year was assigned to that status. The only exception was BTEA, for which any number equal to or above 29 was considered as a full year. This is because BTEA might not be paid over the summer months period between academic years, but this number of weeks would ensure that the individual started attending courses in the new academic year as well, therefore making a yearly coverage reasonable.
7. Weeks of supported employment as identified in the social welfare payments dataset – or in the spell duration for JobsPlus – are attributed as supported employment even in case of lower number of weeks of employment contributions found in the earnings dataset. Total weeks of supported employment are subtracted from weeks of employment contributions recorded in the earnings dataset, to separate periods of supported vs. unsupported employment. Weeks of BTWA are subtracted from contributions for self-employment (S-class contributions); weeks of JobsPlus, BTWFD, PTJA are subtracted from weeks of employment as employees (A-class contributions).
8. Whenever supported employment and employment without support are recorded in the same year, the sum of yearly weeks is capped at 52.

Despite the many allocation rules set in place, the annual nature of the data limits the precision that can be used in properly allocating states over the different points in time. This is one of the main shortcomings to be kept in mind in carrying out the analysis of sequences.

Moreover, it is worth noting that employment data not only do not contain information on exact periods, but they also do not provide details on hours worked. This makes it impossible, e.g., to identify part-time work that might be carried out by jobseekers while still preserving their jobseeker's status. As a consequence, the analysis in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), Chapter 5, likely overestimates the amount of employment without support in the analysis of sequences.

Finally, as already highlighted, the choice of states considered was also largely due to data availability. Most notably, it was impossible to retrieve information on support received from Local Employment Services[3], as well as on many other possibly relevant schemes, for which information was very sparse (e.g., for JobClub).

6.2.4. Additional contextual information is added from other datasets

The lifetime benefit episode data provide information on duration of presence on the Live Register prior to the qualifying eligibility episodes considered in the analysis. As mentioned in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), Chapter 4, this dataset provides individuals' cumulated benefit history prior to the start of their eligibility period, concerning Back to Education Allowance, Live Register, One-parent Family Payment, Tús, and "other" benefits. Information on previous summed duration in registered unemployment (on the Live Register) is very useful to determine whether the individuals in the sample already have a long history of unemployment in the past, and therefore to carry out analyses on individuals with different histories. Eventually, ALMP and work-related information described above is complemented with demographic data including age, gender, marital status, and nationality. On this last source, only one relevant operation has been done, i.e., extending to the full timeframe under scrutiny the pieces of information that reasonably held fixed, i.e., nationality and gender.[4]

6.3. Robustness checks

This section presents some investigation on how robust the sequence analysis is to changes in underlying data aggregations and classifications. These considerations are important in the discussion of sequence analysis, because the choice of data aggregation- both in terms of the length of spells considered and in the definition of a particular state- is fundamental to the interpretation of any results.

6.3.1. Overlapping of states does not frequently occur

As mentioned above, in the analyses described in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), individuals in the data were assumed – for each observed month during the observed timeframe – to belong to only one state, although states could be overlapping. This section explores the potential impact of this assumption made in the analysis.

Once individual histories are broken down month by month and data selection is done as described above, 48 439 296 months appear in the data. Out of these, for 46 895 793 (96.8%) there is no overlapping, as only one state is recorded. For another 3.1% two different states overlap, while in the remaining less than 0.1% of months the overlapping states are three or four. Deeming the latter negligible in terms of potential implications for the results, a focus is made on the 1 522 941 months for which two states are observed at the same time. Two different state combinations explain 64.5% of the two-state overlaps: CE and EWoS (32.7%), and EWoS and JobPath (31.8%). CE and EWoS do not entail any interpretative problem, as they can legitimately come combined provided an individual receiving CE works for a limited amount of time. Indeed, eyeballing of data reveal that such cases appear as long CE episodes with short spells of EWoS within. Overlapping of EWoS and JobPath are instead at least partially due to data construction choices. Around one in four JobPath episodes have missing end date. In these cases, end of episode was arbitrarily set twelve months after the start date. Of course, this may well be at the origin of overlaps with following different states.

Overall, hence, overlaps involve an extremely limited portion of the sample. Out of this, only a limited share is related to data construction choices, which to this respect are hence largely inconsequential. This evidence delivers two reassuring messages. First, different choices in the lexicographic ordering of states are likely not to affect the results or the analyses. Second, although much work is still to be done to have a fully-fledged public use file for research of ALMPs in Ireland, the state of the art of raw data represents a highly reliable starting point.

6.3.2. Casual claims may be classified as a separate state

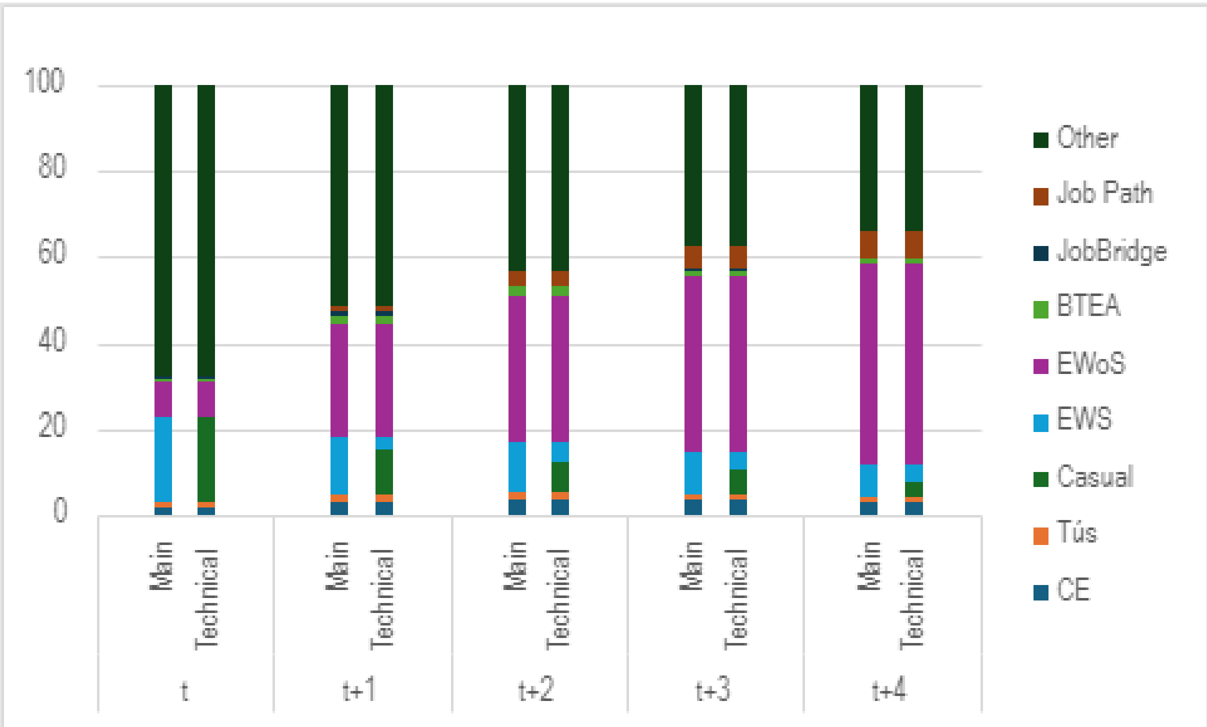
In Chapter 5 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) casual claims are included within the broader category of employment with support (EWS). It might however be interesting to consider this state as a separate one, to understand the relative relevance of this when compared to other forms of employment with support, which are closer to what is normally intended as ALMP participation. In the following, two of the main figures from Chapter 5 of

(OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1])-based on the eight-state lexicographic ordering described above – will be reproduced using casual claims as a state on its own, and imposing the following lexicographic ordering:

1. Community Employment
2. Tús
3. Casual claims
4. Employment with support (EWS)
5. Employment without support (EWoS)
6. Back to Education Allowance (BTEA)
7. Job Bridge
8. Job Path
9. Other

Figure 6.1. Casual claims initially form all of Employment With Support but their share halves over time

Observed state over time for individuals becoming eligible to CE or Tús in 2012-2015



Note: Each vertical bar sums to the 253 578 individuals who became eligible between 2012 and 2015 and is normalised to 100%. They are split proportionally according to the status observed. The bars at the extreme left (labelled t) represents the conditions observed at entry into eligibility, while those on its right represent the states observed one (t+1) to four (t+4) years later. Each time-period shows a comparison between “Main” (Figure 5.4 in the main report, (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) and “Technical” (the present report). BTEA: Back to Education Allowance; CE: Community employment; EWS: Employment with support; EWoS: Employment without support; “Casual”: casual claims. “Other” refers to persons who are not employed nor engaged in any of the supporting schemes covered in the analysis.
Source: Calculations based on Department of Social Protection (DSP) data.

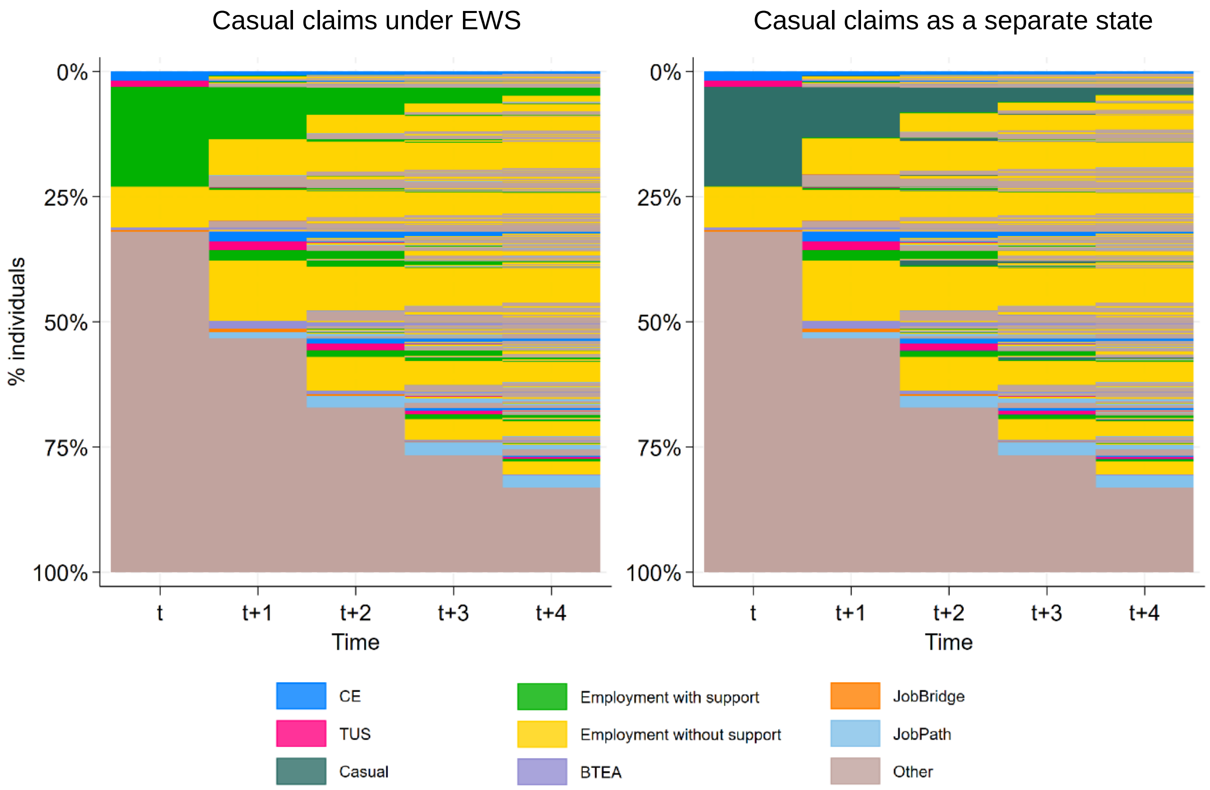
For each time-period from *t* (the moment when a long-term unemployed individual enters eligibility to CE or Tús) to *t+4* (four years later), Figure 6.1. compares the states as classified in Figure 5.4 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) – left bars ("Main") – and states as reclassified here – right bars ("Technical"). Each bar sums to the 253 578 individuals who became eligible between 2012 and 2015 and is normalized to 100%. The introduction of a new state which was included in EWS in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) affects only how EWS is split, while the shares of the other states remain unchanged. At entry, nearly the total of those previously classified as EWS is now observed as a casual claim. While the sum of EWS and casual claims under "Technical" progressively shrinks from 20% to 7.5% in *t+4* – following the same path displayed by EWS only under "Main" vertical bars – the relative share of casual claims within this sum falls to around 50%. In other words, once casual claims are spotted, the initial large share (20%) of EWS turns out to be entirely explained by casual claims. Other types of EWS appear gradually, and grow to 3.7% four years after entry, while casual claims fall to 3.9%.

Looking at individual trajectories,

Figure 6.2. reproduces the analysis shown in Figure 5.8 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]), also keeping casual claims as a separate state. The figure confirms the evidence from Figure 6.1. that upon becoming eligible, most of those previously observed as employed with support are indeed jobseekers under casual claims. Around 23% of individuals in the sample are observed under a casual claim over the time period considered, while only 9% are observed in other forms of employment with support at least once. The vast majority (86%) of those who are found under a casual claim over the time period considered, are already in this state upon becoming eligible for CE or Tús. The opposite holds for other forms of employment with support, as it takes time for jobseekers to start participating in other schemes. 63% of those on a casual claim move directly to being employed without any form of support.

Figure 6.2. Nearly two thirds of those on a casual claim move to employment without support

5-year sequences of states for individuals becoming eligible between 2012 and 2015



Note: BTEA: Back to Education Allowance; CE: Community employment. "Other" refers to persons who are not employed nor engaged in any of the supporting schemes covered in the analysis.

Sequence index plot of the sequences observed in the data for the 253 578 individuals becoming eligible between 2012 and 2015. Sequence index plots use line segments to show how individuals move between states over time. Changes of colour represent changes in state. The Y axis shows one line per individual. The X axis represents time points from the moment individuals become eligible; t represents the conditions observed when becoming eligible, while the other points represent the states observed one (t+1) to four (t+4) years later. The left panel shows the sequence index plot as presented in Figure 5.8 in the main report (OECD, 2024), with casual claims included under EWS; the right panel shows casual claims as a separate state.
Source: Calculations based on Department of Social Protection (DSP) data.

6.3.3. Sequences with data at monthly frequency reiterate the main messages

The analysis presented in Chapter 5 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) on the sequences of programmes in which individuals participate was based on a longitudinal dataset where individuals were observed at yearly intervals starting from the moment in which eligibility started. As described in the previous sections, the underlying dataset was however built with a monthly structure, which in theory would allow depicting individual trajectories in a more detailed way. As explained in the Country Report, the choice of using yearly intervals was made to retain a strategic view of the main patterns observed, while reducing the potential number of sequences observed in the data; ultimately, this selection was made to facilitate interpretation of results.

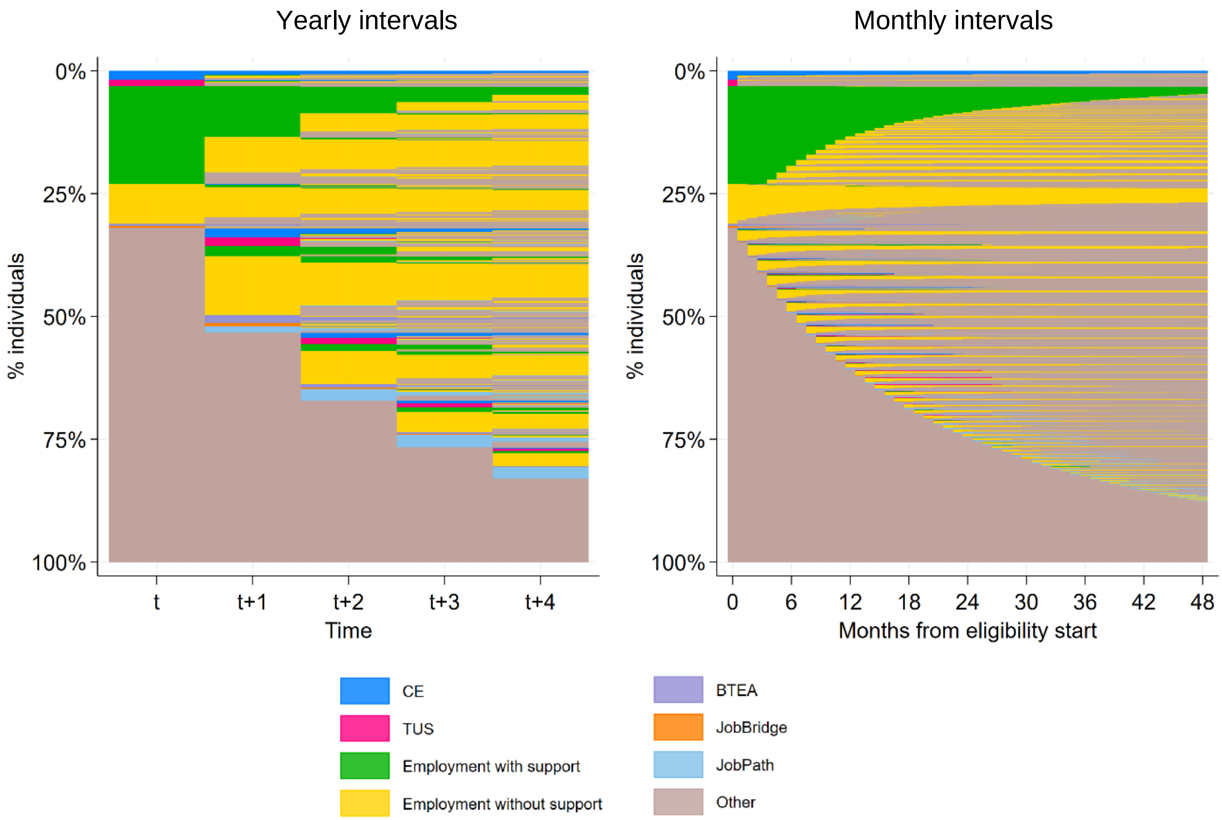
The downside of this choice is the loss of information on short-term movements that might happen within the year. Figure 6.3. shows a comparison between results from Figure 5.8 in the Country Report, based on the use of yearly intervals (left panel), and the same type of analysis done exploiting the full monthly structure of the dataset (right panel). As expected, the use of high-frequency intervals makes the results somewhat more difficult to interpret, as individual trajectories become more fragmented, and less common sequences are less visible. On the other hand, the figure shows even more clearly some of the main messages emerging from the analysis at yearly intervals, namely:

- The high share of individual trajectories starting while being employed with some form of support;
- The likelihood of transition from this state to employment without support;
- The equally large share of individuals never observed participating in any of the schemes considered in the analysis.

Overall, this section shows that despite the potential loss of information due to the choice of yearly intervals in the analysis, the main conclusions that were drawn from the analysis are not undermined, but even reinforced, when exploiting the entire monthly structure of the dataset.

Figure 6.3. The analysis of monthly sequences instead of yearly ones does not change the main findings

5-year sequences of states for individuals becoming eligible between 2012 and 2015



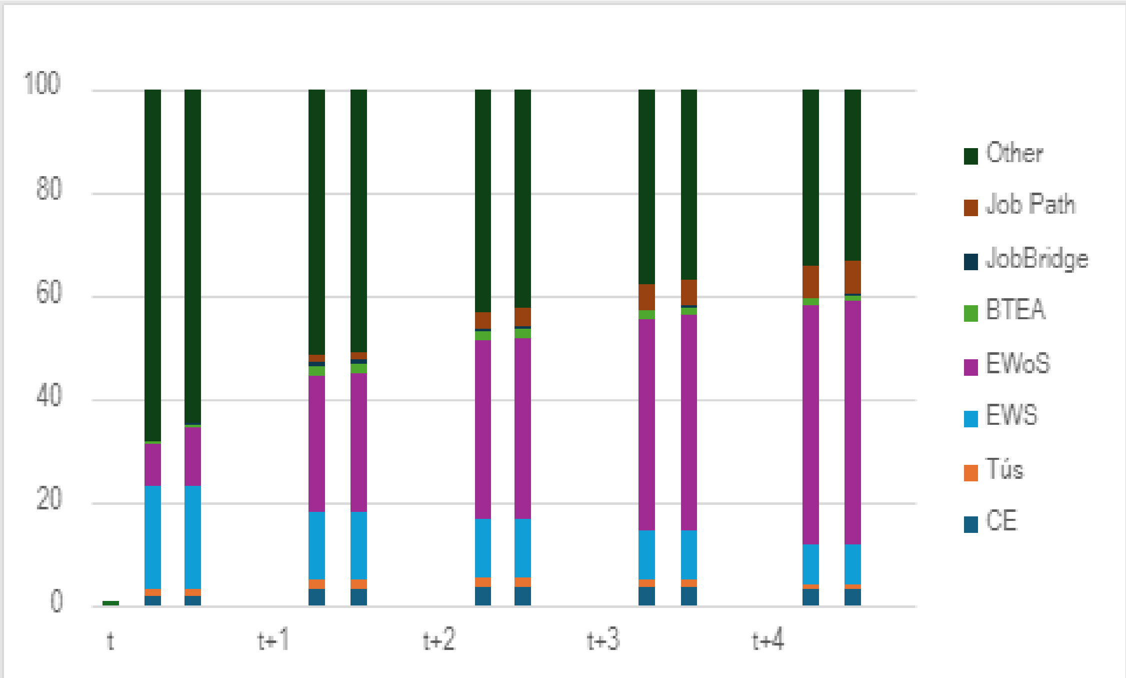
Note: BTEA: Back to Education Allowance; CE: Community employment. "Other" refers to persons who are not employed nor engaged in any of the supporting schemes covered in the analysis.
Sequence index plot of the sequences observed in the data for the 253 578 individuals becoming eligible between 2012 and 2015. Sequence index plots use line segments to show how individuals move between states over time. Changes of colour represent changes in state. The Y axis shows one line per individual. The X axis represents time points from the moment individuals become eligible. The left panel shows sequences with individuals observed at yearly intervals, as presented in Figure 5.8 in the main report (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]); *t* represents the conditions observed when becoming eligible, while the other points represent the states observed one (*t+1*) to four (*t+4*) years later. The right panel shows sequences with individuals observed at monthly intervals.
Source: Calculations based on Department of Social Protection (DSP) data.

6.3.4. Changing the assumptions on EWS and EWoS temporal distribution does not affect the results

As described in more details above, much of the raw data feeding the information on EWS and EWoS does not allow a clear allocation of the observed weeks over a certain year, because records are only available at aggregate annual level. Start and end dates therefore had to be imputed. Most importantly, when the number of credited weeks is lower than a year and there is clear evidence that they do not span over consecutive years, (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) assumes that EWS comes first (i.e. starting from January), and EWoS last (going backwards from December). This assumption may overstate the frequency of transitions from EWS to EWoS. As a robustness, Figure 6.4. compares the evidence from Figure 5.4 in (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) with the same computations under the opposite assumption that in such undecided cases EWoS precedes EWS. Data is again sampled to retain individuals entering eligibility in 2012-2015, and Figure 6.4. reads as Figure 6.1. . As clear, although present, changes are negligible. Neither the strongest assumption made on start and end dates of EWS and EWoS is hence able to significantly affect the results in Chapter 5 of (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]).

Figure 6.4. Changing the assumptions behind EWS and EWoS start and end dates is inconsequential

Observed state over time for individuals becoming eligible to CE or Tús in 2012-2015



Note: Each vertical bar sums to the 253 578 individuals who became eligible between 2012 and 2015 and is normalised to 100%. They are split proportionally according to the status observed. The bars at the extreme left (labelled t) represents the conditions observed at entry into eligibility, while those on its right represent the states observed one (t+1) to four (t+4) years later. Each time-period shows a comparison between "Main" (Figure 5.4 in the main report, (OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre, 2024[1]) and "Technical" (the present report). BTEA: Back to Education Allowance; CE: Community employment; EWS: Employment with support; EWoS: Employment without support. "Other" refers to persons who are not employed nor engaged in any of the supporting schemes covered in the analysis.
Source: Calculations based on Department of Social Protection (DSP) data.

References

A. Smith, J. and P. E. Todd (2005), "Does matching overcome LaLonde's critique of nonexperimental estimators?", *Journal of Econometrics*, Vol. 125/1-2, pp. 305-353, https://doi.org/10.1016/j.jeconom.2004.04.011. [17]

Abadie, A. and G. Imbens (2016), "Matching on the Estimated Propensity Score", *Econometrica*, Vol. 84/2, pp. 781-807, https://doi.org/10.3982/ecta11293. [37]

Abadie, A. and G. Imbens (2008), "On the Failure of the Bootstrap for Matching Estimators", [36] *Econometrica*, Vol. 76/6, pp. 1537-1557.

Anders, J. and R. Dorsett (2017), “What young English people do once they reach school-leaving age: A cross-cohort comparison for the last 30 years.”, *Longitudinal and Life Course Studies*. [30]

Busso, M., J. DiNardo and J. McCrary (2014), “New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators”, *Review of Economics and Statistics*, Vol. 96/5, pp. 885-897, https://doi.org/10.1162/rest_a_00431. [15]

Caliendo, M. and S. Kopeinig (2008), “Some practical guidance for the implementation of propensity score matching”, *J Econ Surv*, Vol. 22/1, pp. 31-72, https://doi.org/10.1111/j.1467-6419.2007.00527.x. [26]

Caliendo, M., R. Mahlstedt and O. Mitnik (2017), “Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labor market policies”, *Labour Economics*, Vol. 46, pp. 14-25, https://doi.org/10.1016/j.labeco.2017.02.001. [11]

Card, D., J. Kluve and A. Weber (2018), “What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations”, *Journal of the European Economic Association*, Vol. 16/3, pp. 894-931, https://doi.org/10.1093/JEEA/JVX028. [18]

Card, D., J. Kluve and A. Weber (2018), “What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations”, *Journal of the European Economic Association*, Vol. 16/3, pp. 894–931, https://doi.org/10.1093/jeea/jvx028. [12]

Dengler, K. (2019), “Effectiveness of sequences of classroom training for welfare recipients: what works best in West Germany?”, *Applied Economics*, Vol. 51/1, pp. 1-46. [23]

Dengler, K. (2015), “Effectiveness of sequences of One-Euro-Jobs for welfare recipients in Germany,”, *Applied Economics* 6170-6190, pp. 47:57, https://doi.org/10.1080/00036846.2015.1066488. [22]

DSP (2023), *Pay Related Social Insurance (PRSI) Contribution Rates and User Guide*, https://assets.gov.ie/216225/98adc7d5-8641-4e9d-8f9c-5907adecdd2a.pdf. [4]

DSP (2021), *gov.ie/tusoperationalguidelines*. [29]

Frölich, M. (2004), “Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators”, *Review of Economics and Statistics*, Vol. 86/1, pp. 77-90, https://doi.org/10.1162/003465304323023697. [14]

Government of Ireland (2023), *Operational Guidelines: Tús – Community work placement initiative*, https://www.gov.ie/en/publication/e67506-operational-guidelines-tus-community-work-placement-initiative/. [2]

Government of Ireland (2023), *The different classes of Pay-Related Social Insurance (PRSI)*, https://www.gov.ie/en/publication/14ecbe-the-different-classes-of-pay-related-social-insurance-prsi/. [3]

Halpin, B. (2013), “Imputing Sequence Data: Extensions to initial and terminal gaps, Stata’s mi *”, *University of Limerick, Department of Sociology Working Paper Series*. [31]

Heckman, J. et al. (1998), “Characterizing Selection Bias Using Experimental Data”, [9] *Econometrica*, Vol. 66/5, p. 1017, https://doi.org/10.2307/2999630.

Huber, M., M. Lechner and C. Wunsch (2013), "The performance of estimators based on the propensity score", *Journal of Econometrics*, Vol. 175/1, pp. 1-21, https://doi.org/10.1016/j.jeconom.2012.11.006. [13]

Imbens, G. (2000), "The role of the propensity score in estimating dose-response functions", *Biometrika*, Vol. 87/3, pp. 706-710, https://doi.org/10.1093/biomet/87.3.706. [16]

Kelly, E., S. McGuiness and J. Walsh (2015), *An evaluation of the Back to Education Allowance.*, ERSI and DEASP. [28]

Lechner, M. (2012), "Sequential Causal Models for the Evaluation of Labor Market Programs", *http://dx.doi.org/10.1198/jbes.2009.0006*, Vol. 27/1, pp. 71-83, https://doi.org/10.1198/JBES.2009.0006. [24]

Lechner, M. (2004), "Sequential Matching Estimation of Dynamic Causal Models", *University of St. Gallen Department of Economics working paper series 2004*, https://ideas.repec.org/p/usg/dp2004/2004-06.html (accessed on 23 February 2023). [21]

Lechner, M. and R. Miquel (2001), "A Potential Outcome Approach to Dynamic Programme Evaluation: Nonparametric Identification", *SSRN Electronic Journal*, https://doi.org/10.2139/SSRN.274825. [19]

Lechner, M. and S. Wiehler (2013), "Does the Order and Timing of Active Labour Market Programmes Matter?", *Oxford Bulletin of Economics and Statistics*, Vol. 75/2, pp. 180-212, https://doi.org/10.1111/OBES.12007. [25]

Lechner, M. and C. Wunsch (2013), "Sensitivity of matching-based program evaluations to the availability of control variables", *Labour Economics*, Vol. 21, pp. 111-121, https://doi.org/10.1016/J.LABECO.2013.01.004. [10]

Ministry of Social Security and Labour (2017), *Įsakymas dėl užimtumo rėmimo priemonių įgyvendinimo sąlygų ir tvarkos aprašo patvirtinimo [Order on the approval of the conditions and procedure for the implementation of employment support measures]*, Republic of Lithuania. [35]

OECD (2023), *Evaluation of active labour market policies in Finland*, OECD Publishing, Paris. [32]

OECD (2022), *Impact Evaluation of Vocational Training and Employment Subsidies for the Unemployed in Lithuania. Connecting People with Jobs.*, OECD Publishing, Paris, https://doi.org/10.1787/c22d68b3-en. [33]

OECD (2021), *Building inclusive labour markets: Active labour market policies for the most vulnerable groups*. [27]

OECD (2020), *Impact Evaluations Framework for the Spanish Ministry of Labour and Social Economy and Ministry of Inclusion, Social Security and Migrations*, https://www.oecd.org/els/emp/Impact_Evaluations_Framework.pdf. [5]

OECD/Department of Social Protection, Ireland/European Commission, Joint Research Centre (2024), *Impact Evaluation of Ireland's Active Labour Market Policies*, OECD Publishing, Paris, https://doi.org/10.1787/ec67dff2-en. [1]

Republic of Lithuania (2016), *Law on employment*. [34]

Robins, J. (1986), “A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect”, *Mathematical Modelling*, Vol. 7/9-12, pp. 1393-1512, https://doi.org/10.1016/0270-0255(86)90088-6. [20]

Sianesi, B. (2004), “An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s”, *Review of Economics and Statistics*, Vol. 86/1, pp. 133-155. [8]

Sianesi, B. (2004), “An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s”, *The Review of Economics and Statistics*, Vol. 86/1, pp. 133-155, https://doi.org/10.1162/003465304323023723. [7]

Sianesi, B. (2008,), “Differential effects of active labour market programs for the unemployed”, *Labour Economics*, Vol. 15/3, pp. 370-399. [6]

Notes

[1] Overlapping of different states includes cases which in principle should not be observed in the data, as not allowed by the law. For instance, we have cases of individuals eligible to CE and also working without support. It is impossible, at the current state of the data, to say why. Maybe some forms of support are not captured in the data; also, it might be the simplification rules used to allocate weeks over the year when only aggregate annual information is available. Eventually, these individuals become employed right after entering eligibility.

[2] At this step eligibility spells are not restricted to the target time-frame, as an eligible CE episode, started before 2012 but ending after 2012 (hence in target), would be so based on an eligibility period outside of the target time-frame.

[3] A referral dataset showed information on LES, which however only concerned referrals to trainings. Such information was only available in a very unstructured way, and with no detailed information on length and intensity of such trainings. This prevented the authors from using these data in the analysis.

[4] Marital status can obviously change over time and is updated when the individual makes a new access to the employment services – which can potentially be even years after the actual change occurred. Age is indeed created through the date of birth and is hence a time-varying variable with reasonably no error.