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

Technical Report: Evaluation of active labour market policies in Finland

February 2023

Table of contents

FIGURES

TABLES

Executive summary

This report is undertaken 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, the OECD conducted a counterfactual impact evaluation (CIE) of labour market training (LMT) and self-motivated training (SMT) in Finland, publishing the results and the consequent policy recommendations in the OECD publication series *Connecting People with Jobs* (OECD, 2023[1]). This technical report accompanies the report on evaluation results, providing more detail on the data and techniques underpinning the main report and more context on some of their main strengths and weaknesses. It includes discussion of the different data sources available for analysis, the process to link and prepare data for analysis, the metadata documentation and the selection of the econometric techniques and how these were applied in the analysis conducted by the OECD. A number of robustness checks that were undertaken to provide context on the strengths and uncertainties of the findings discussed in the main report are also reported here.

In chapter 3 of the main report, the OECD (2023[1]) covered in detail the system of evidence generation that Finland has in place to evaluate its active labour market policies (ALMPs) and this included a set of recommendations to improve the capacity of Finland to continue building this evidence in the future. For this reason, this accompanying technical report does not cover these issues, except for where they relate to data documentation and availability specifically to support such activities.

Finland has a rich array of detailed and high-quality individual level administrative data to draw upon for evaluation of ALMPs. These data are held in a secure data repository at Statistics Finland and are available for use by approved researchers, using secure access protocols which prevent the unauthorised use of data and ensures compliance with legal requirements related to data access. Linking data from different registers was facilitated by the existence of secure, pseudo-anonymised personal identifiers, which are based on underlying social security numbers. Their existence made it extremely easy for datasets with different information to be linked together. These identifiers were available in all the analytical datasets.

Comprehensive panel data, with a long history dating back some 30 years, facilitated a detailed analysis based on rich personal characteristics of jobseekers. These data include age, education, marital status, housing tenure type, presence of children in family, geography, in addition to detailed labour market outcomes (including employment, unemployment, earnings, months worked and occupation) and ALMP participation. The long time series available meant it was possible to construct employment and unemployment histories of individuals prior to their participation in ALMPs, vital to ensuring that participants in ALMPs were compared to non-participants that are similar to them across these past labour market experiences. Some interpretation and guidance from experts on the use of data was required due to the

[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).

existence of different datasets containing programme information which provided different summaries of the individuals participating in those programmes.

The methodology used in the OECD (2023[1]) impact evaluation was determined by the availability of rich administrative data and by lack of any trial evidence of strict eligibility criteria for entry into training. In both LMT and SMT, participation is not randomly assigned and is open to all jobseekers, with only some minor restrictions on lower age limits (individuals under 20 are rarely admitted to LMT and SMT has a lower age limit of 25). In addition to this, it was not possible to use counsellor level data to determine whether some counsellors are better or worse at signposting their clients to these programmes, to investigate whether this might have influenced uptake and therefore outcomes. Due to these issues, an econometric strategy was adopted that uses the rich individual characteristics available to compare participants to similar non-participants. This relies on removing all possible differences using these observable data, so that they are compared as if the participants had been randomly selected into the programmes. In addition, where cohorts are compared across years, this strategy is combined with an estimator that looks only at relative changes over time, to try as far as possible to remove all unobservable differences between cohorts.

To help support the extensive data that are already available for research in Finland, a number of small enhancements to metadata and data availability could be made.

Enhancements to metadata:

- Statistics Finland, or the authorities providing the data to them, to provide better guidance on which variables should be used in which circumstances, particularly where data provide different estimates of the number of individuals affected.
- Improve research consistency by creating a secure area for researchers to share code, particularly relating to data cleaning and foster an ecosystem that promotes these activities (for example, by marketing this area and encouraging its use). SF staff could also write some of these data cleaning routines.
- Ensure metadata on the SF secure data server is up-to-date and consistent with the latest metadata published on the Taika data catalogue. Ensure metadata contains all decode values that are used in datasets.
- Review metadata to include more detailed information on the robustness of individual variables. Particularly where variables have missing values or other variables exist which provide similar but different information.
- Review the provision of English metadata for non-native researchers, to minimise analytical re-work and the risk of incorrect translations.

Enhancement in data availability:

- Include data which enables linking job counsellors to jobseekers and which includes details on meetings between the two sets of parties.
- Improving the timeliness of data uploads would facilitate more up-to-date data analysis. Data which provide more temporal disaggregation would allow short term labour market dynamics to be studied. For example, real-time earnings data would both facilitate much more timely evaluation and provide insight into short-term labour market outcomes.
- Data on jobseeker interactions with digital administrative systems could provide useful insight into jobseeker behaviour and engagement and could facilitate improvements to how counsellors use digital tools to support jobseekers. This could also be especially helpful for the Job Market online platform, to facilitate better matching tools between employers and jobseekers.
- Better coverage and data content on training could further improve insight into their functioning.
- Including data on wider impacts, such as on health and crime, could enable broader ALMP benefits to be considered.

1 Data content and documentation

This chapter details the data available for use in the evaluation of labour market training (LMT) and self-motivated training (SMT). These data were accessed using the Statistics Finland online data repository. This is a secure virtual server which allows approved researchers tightly controlled and monitored access for specific, agreed research purposes. It allows access to a rich source of individual level information which is essential to the conduct of the impact evaluation conducted by the OECD (2023[1]). The chapter highlights some of the advantages of these data, alongside some of their limitations and areas where improvements would help researchers and policy makers to generate an even better evidence base than is currently the case.

1.1. Data access and time span

The data used in the evaluation are high quality individual level administrative data. These data cover a long time period and a large number of personal characteristics which facilitate the type of quasi-experimental CIE that is conducted by the OECD.

1.1.1. Data access is organised through Statistics Finland

The data access process is controlled via a centralised application process at Statistics Finland (SF). Researchers use a standardised application template which enables applications to be assessed by staff on a consistent basis. Within this process, researchers have to summarise the objectives and main content of their proposed research project, list their research questions and outline their data needs. Once this has been agreed by SF, researchers individually sign a pledge of secrecy to ensure that they uphold Finnish law in accordance of their use of the data and do not disclose any data unlawfully.

SF has around 60 data collections which cover a vast scope of dimensions on individuals and firms within the Finnish economy. These data collections can be categorised into 12 higher-level “families” of data. The two families of data that are relevant to the evaluation conducted in this report are:

- FOLK – which comprises 12 sub-datasets and covers a large range of derived individual level datasets, including those data drawn from other families in the broader SF collection (for example FOLK dataset include information derived directly from TEM administrative data).
- TEM – collections of data ingested directly from the Ministry of Economic Affairs and Employment (TEM) covering all aspects of the labour market services provided to jobseekers. Organised as four data collections, with the main “job search” data collection organised further into 11 sub datasets.

1.1.2. The key variables cover a range of personal characteristics over a long time period

The data available at SF are rich and cover a long time period (Table 1.1). The FOLK basic data, which are used to derive the sample of individuals unemployed at the end of the year and their corresponding personal socio-demographic characteristics are available going back to 1987 and run until 2021 (though the analysis conducted for this project used the previous annual data covering the period to 2020). TEM administrative data go back to 1991. However, despite this headline data availability, many variables within this dataset are available with a one- or two-year lag to the specified time period. For example, the variable used for annual earnings in the quantitative analysis in this report is available only until two years prior to the dataset date. The real-time register data in Finland has already been used for research purposes and legislative reform will bring it into use at KEHA and TEM. Making these data available to the wider research community would bring significant benefits to the timeliness of evaluations and their ability to analyse employment trajectories.

The FOLK basic data provide a wealth of annualised information that covers an individuals' personal socio-economic characteristics. It includes their sex, age, geographical location, marital and family status, their level and subject area of education, their occupation, the numbers of months of unemployment and employment in the year, their housing situation, the presence of children in the family, alongside a number of different categorisations of annual income.

In addition to these data, other FOLK data are used to provide information on additional dimensions. The FOLK income dataset allows the presence of the amount of annual study subsidy (a grant that contributes towards continuing study costs) to be identified. The FOLK unemployment dataset records all spells of unemployment. This allows a calculation to be performed which summarises the number of unemployment spells in a year and the cumulative days of duration of unemployment over this period.

The TEM job search data provide detailed information on all aspects of labour market services provided to jobseekers. To facilitate quicker data processing the overall data summarised in Table 1.1 for TEM job search are organised into 11 separate underlying datasets. These are: Person's employment data, persons job search data, person's intervention placements, work experience at the workplace, labour market training data for a person, target occupation for labour market training, job placements for a person, statements of a person's employment policy, mobility allowances, person's job search plans, labour market training course details. For the OECD (2023[1]) analysis the labour market training dataset is used to provide start and end dates of LMT spells and the statements of a person's employment policy dataset provides similar information for the SMT spells. Unlike the data for LMT, where TEM has good administrative records of the precise course that the individual started and the participation in that course, the data for SMT are more diffuse. They do not link directly to SMT course information and were first implemented after the introduction of SMT in 2010 to monitor uptake of SMT. Rather they record an intent (and agreement from the TE office) that the individual wishes to and can start the training, but they do not link to actual participation. For this step to be taken, one must link the SMT information from TEM to data on educational attainment from the Ministry of Education and Culture (OKM).

Table 1.1. Individual micro-data have a long time series and are comprehensive

Data collections available at Statistics Finland used for the OECD analysis

Datasets	Time Period	Variables
FOLK Basic	1987-2020	Year, Encrypted SF's identity code, Year of birth, Year of death, Sex, Region of birth, Municipality of residence at birth, Country of birth, Nationality, Origin, Language, Person's age, Region of residence, Municipality, Statistical municipality group of municipality of residence, Municipality (statistical year), Statistical municipality group of municipality of residence (Dec 31st), Urban-rural classification, Place of residence in an urban settlement, Marital status, Socio-economic group, Matriculation examination, Educational level of highest qualification/degree, Educational field of highest qualification/degree, Year when the highest qualification/degree was attained, Occupational status (TVM), Type of education, Received student financial aid, Main type of activity (TVM), Code of occupation, Number of unemployment months, Months in employment, Size of household-dwelling unit, Type of household-dwelling unit, Tenure status of dwelling, Number of rooms (kitchen not included), Type of building, Standard of equipment, Age of youngest child in family, Family status, Number of children in family, Number of all children aged under 18 in the family, Number of children aged under seven in the family, Number of children aged under three in the family, Size of family, Family type, Old-age pension, Disability pension, Unemployment pension, Special pension for farmers, Part-time pension, Survivor's pension,

Datasets	Time Period	Variables
		Individual early pension, Disposable income, Debts in total, Taxable assets, Earned income total in state taxation, Earned income, Entrepreneurial income (own), Car ownership,
TEM job search	1991-2020	Protected TK personal identification number, Year of birth, gender, Job search start date, Date of termination of previous employment, Country of transmission, Reason for ending the job search, Pre - employment activities, Job search end date, Residence / work permit, Industry of previous employer-2, Reason for termination of previous employment, Applicant from abroad, Start date, End date, Reason for change in employment, Employment code, Start date, Employment profession, ESF project code, Target group, Term/end date, Type of placement subsidy funds, Working hours, Staff size class, Type of employment, Reason for termination of employment, Employer's business and company identification number, Sectors, Employer's industry 2 (missing start-up funds), Type of plan, Date of plan, start date, Unemployment Insurance Fund number, KELA's office number, Opinion code, Period end date, date of opinion, Date of commencement of employment, Brokerage profession, Job advertisement number, final date of termination of employment, Forwarding TE centre, Employment area, Date of arbitration, Workplace Support, Start date of the work trial, ESF project code, Duration of the experiment (in working days), Publisher of the experiment, Start code, Occupation of the experiment, Type of experiment, Objective of the experiment, Sector of experiment, Start code, Start date of training, Occupation of placement, Date of application for training, Pre-training activity, Type of training, Number of days of training, Course identification number, City of residence, Quality of employment relationship, Date of end of training, Reason for termination, Type of placement, Employment code, Employment, Selection decision, Code of the ELY centre providing the training, Special group, ESF project code, Course start date, TRAINING, Course hour number, Type of training, Training association, Type of course, Main language of instruction, Course cancelled, Group course, Occupation detail, Person's previous target occupation, Date of change, type of aid, target centre, grant date, Type of benefit, Apprenticeship, Protected organiser code, Proportional share of aid paid, State-funded education (from 2018),
FOLK unemployment	1987-2019	Statistical year, Encrypted SF's identity code, Start date, End date, Reason for ending, Unemployment Code
FOLK Income	1987-2019	Year, Encrypted SF's identity code, Earned income, Entrepreneurial income, Business income, Earned income from agriculture, Income subject to state taxation, Earned income total in state taxation, Capital income total in state taxation, Current transfers received, Sickness allowance, Pension income, Child home care allowance and partial care allowance, Parental allowance, Unemployment benefits, Housing benefits, Study grant, Social assistance, Daily and maternity allowance, Current transfers paid, State income tax, Share of income tax of earnings taxes, Share of income tax of capital gains taxes, Municipal tax, Church tax, Property tax, Disposable money income, Taxable assets, Debts in total

Note: Variables used in the main OECD country report (2023[1]). TEM variables are machine-translated and no human editing has taken place.
Source: Statistics Finland metadata catalogue and OECD translations

1.2. Limitations of existing data

Despite the excellent individual level data that are available in SF, there are several areas where data availability could be improved to add further power and insight to policy analysis:

- Linked counsellor-jobseeker data are not available – there are no data on the counsellors that are linked with data on jobseekers. This is an important source of information, not only for analysing the interactions between counsellors and jobseekers themselves, but also offering a possible source of variation for conducting broader policy analysis. For example, if counsellors vary in how they refer individuals to different programmes, then it is possible to look at the outcomes for jobseekers by counsellor to see whether they influence individuals' outcomes.
- Information on job counselling meetings between jobseekers and counsellors is not available – information on the specific interactions counsellors have with the clients can provide insight on how these meetings help individuals to find work. These data can take various forms, from a basic record of the meeting date to including information on things like counsellor referrals to jobs and the duration of the meeting. A recent paper in Sweden used such data to look at not only the direct effects of job search interviews, but also associated indirect effects via displacement of other

jobseekers (Cheung et al., 2019[2]). Consideration of the reporting of these data is now doubly important in Finland because of the two reforms to its labour market services that it has embarked upon. The reform to its customer service model introduces fortnightly meetings between counsellors and jobseekers and is predicated on savings to the exchequer due to quicker jobseeker employment. There is currently no specific evidence from Finland to corroborate these assumptions (which are based on reforms in other countries). Secondly, the transfer of powers to municipalities to deliver labour market services means that there is a potential for coordination of data reporting to become harder.

- Information on digital transactions is missing– Jobseekers are primarily directed to apply for employment services (TE) online. Labour market training courses and applications are also conducted online. Utilising data on how jobseekers interact with these digital channels may provide evidence across different dimensions. However, such data are currently not available in SF. 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. Many countries already use such profiling services, but they tend only to use questionnaires or traditional administrative data on benefit receipt (Desiere, Langenbucher and Struyven, 2019[3]), rather than observing individuals' behaviour interacting with services. This may also be of direct use in the Job Market online platform, to facilitate innovation in the matching of employer vacancies and jobseekers. Similarly collecting data on such use then also offers the possibility for quick and easy randomised testing of such online menus, programmes and websites, because it easy at a large-scale to randomise what such menus and screens a jobseeker sees. This may be to alter information provision or to change how menus are processed. This opens up the possibility for innovation. A recent example of a digital tool aimed at helping caseworkers refer jobseekers to appropriate vacancies and ALMPs is an example of using digital technology to do such work (OECD, 2022[4]).
- Short-term earnings data are not available in the SF catalogue– having only annualised data on earnings means it is impossible to look at how ALMPs influence short term labour market transitions. This is especially important for those employment measures and services that are shorter-term in nature, for example like job counselling. If the impacts on helping people into work are small, for example in the time taken to search for work and impacts on earnings in that work, then it is unlikely these impacts will be detectable in annualised earnings, even if they do exist within the year. More disaggregated earnings data can overcome some of these issues and enable these shorter-term dynamics to be investigated (For an example of where short-term employment dynamics are explored following job-search interview see DWP (2018[5])). Whilst the same issues are of less salience for programmes that are expected to have longer-term benefits, such as training, they can also help to give more insight into dynamics of "lock-in" periods, whereby individuals delay their job search whilst they participate in programme. The Finnish Tax Administration does have an almost real-time income register (lagged by one month), which is available for research purposes. The easy incorporation of these data into Statistics Finland (either routinely or ad hoc) would enhance the possibilities research using that gateway.
- Data aggregations can miss detail – most FOLK datasets record information only at the end of each calendar year. These data then miss potential transitions within year. For example, if an individual changes occupation in the year, only the occupation at the end of the year, or the occupation for which the individual had the longest period of time in will show in the data.
- Limited information on training – both LMT and SMT lack contextual information. For LMT there is no information available on the type of course studied, for example the field of study, or the precise training type. Such data would permit an investigation into whether some particular types of LMT are better than others. For SMT no data exists specifically from TEM, other than proposed start and end dates of training. It is possible to link this information to data held on educational

attainment, however this comes with some coverage and interpretation problems. For example, around 12% of individuals first instance of SMT in the data cannot be matched with an education record in that same year. Another 6% of individuals are associated with a course (that is not vocational training) that directly continues the same study in the same or previous year, which should not occur unless studies started during employment or when receiving employer paid financial benefit.

- No health or justice data – to look at wider impacts of ALMPs, for example how the participation in an ALMP affect usage of health services or criminal misdemeanours.
- Some FOLK spells datasets structure requires processing and is not immediately conducive to analysis – records are defined annually and do not constitute individual spells (these datasets are different to the annual FOLK datasets, where each individual has only one record per year, for example the FOLK basic dataset used to define unemployment at the end of the year). For example, a spell of unemployment that spanned two calendar years would have two annual records, the chronologically earlier record with a missing end date and the second record with the end date recorded. However, a year containing two separate spells would have two records for those spells. So the data are neither uniquely annual (one row per year) nor uniquely spell based (one row per spell). The adjustments are simple enough to make to transform the data into one or the other, as long as the data generation rules are known (which are not given in the metadata), but it introduces unnecessary complexity for researchers and the possibility for the incorrect processing of data.
- Data are not particularly timely – this is not a problem for impact analysis per se, but it does limit the ability for evidence to be built quickly for programmes. For example, the most up-to-date earnings information at present is for 2019. This kind of speed could be essential when needing to quickly scale-up recently implemented policies in times of acute labour market pressure. For example, there was a reform implemented in 2022 to labour market services which increased activation meetings between counsellors and jobseekers. If a large negative labour market shock occurred and the government wanted evidence on whether to increase recruitment to conduct more interviews, at present it would not be able to generate any post-hoc evidence that its current policy offers value-for-money before doing so.

1.3. Metadata

SF offers a metadata catalogue for its datasets that is comprehensive and generally well documented, but it can lack key information on dataset production and English language translations can be incomplete.

1.3.1. Metadata provide source and variable information across datasets

The SF metadata catalogue offers researchers a good overview of the datasets that are held in SF and their key characteristics. Some of the key features of the metadata are the following:

- Documentation exists for all datasets – there is a metadata entry for every dataset that SF gives researchers access to.
- Defines sample – generally each metadata file provides information on the sample of individuals that are eligible for inclusion into the dataset.
- Defines timeframe – information is given on the coverage period of the data. This information often details particular features (definition of variables/increased or reduced variable coverage) that have changed over the course of the datasets.
- Provides identifying information on the publication date for the metadata and on the underlying dataset names. This is helpful for ensuring that researchers can identify how up to information is.

For example, dataset names often have the coverage years as a suffix to them, so the researcher is able to identify for which annual vintage the metadata relate to.

- Provides a list of all variables in short and long format – the short format list simply provides the name and a variable description, which aids in a quick visualisation of the dataset content. This then links down to a long list which provides variable short and long names, the grouping that the variable belongs to and decode values.

On top of higher-level documentation available for datasets, there is some good practice on additional documentation to outline the hierarchy of some of the derived variables in a dataset. For example, in the FOLK basic data, a PowerPoint presentation is available that outlines in a stepwise process how the variable "ptoim1" is derived. This is a hierarchical variable which defines the labour market status of an individual at the end of the year and so the categorisation of an individual into a particular group may depend on the existence of conflicting pieces of information. The hierarchy enables researchers to determine what precedence these pieces of information are given.

1.3.2. There are a number of metadata areas that could be improved

Despite the fact that metadata do exist and cover the datasets that are available to researchers, there are a number of areas in which they could be improved further which would both reduce burdens to researchers and SF staff and help to increase consistent between separate analytical projects.

Metadata on the secure SF server are often outdated and are hard to process

Although more recent metadata are available online, at the SF Taika research catalogue[1], in many cases the documents available directly on the server where the data analysis is conducted are out-of-date. For example, the TEM metadata, which are available in Finnish only, contain information on the datasets up until 2016 only. These metadata are available on the Taika website on the internet until the year 2021. This poses two problems. The first is that researchers might consider the information contained on the secure server as the most up-to-date source of information and may therefore not check whether more exists on the Taika catalogue, which is outside the secure environment. They then will either potentially misinterpret variables, omit them completely, or increase the workload to SF analytical staff by sending questions to them to answer. The second problem relates to analytical efficiency. If researchers want to load decode values into Stata and have located the up-to-date metadata via the Taika catalogue, they would either have to manually transcribe these values, or ask for the file to be uploaded by SF staff to the secure server. Both solutions are more inefficient than if these data were already present on the secure server.

Having the metadata already uploaded to the server in the form of look-up tables would enable researchers to much more quickly load formats into their analysis, so that it is quick and easier to interpret datasets. The cost of physical storage space requirements of such additions is minimal and with some kind of mark-down coding structure, it should be possible to automate the process so that when metadata documentation is updated, it could be easily uploaded to the server. This could also help to rectify the previous point on the out-of-date metadata on the SF secure server.

Metadata sometimes do not have full variable decodes

Even the metadata that does exist is not always comprehensive in the variable decodes it contains. For example, when looking at the labour market statement codes (variable "lausno") on the TEM data, several of the values are not present anywhere in the metadata. These may comprise only a small minority of the overall data, however, because these codes have no text descriptors anywhere, it makes it impossible for researchers to properly account for these data in their analysis

English metadata are not available for all datasets and where they are, decode values are not always present

A number of metadata for different datasets have not been translated into English. This includes all of the TEM data. For non-Finish speaking researchers, this significantly increases the analytical complexity. In the first instance, it means that translation services are required. Freely available translation services make this task somewhat easier, however there are still issues involved. The Finnish metadata versions often have abbreviated text decode values. Therefore, translation services have to interpret these. For many values, this can be done without much of a problem, however, there remain a minority for which this is not the case. As an example, for the same decode value, two different translation services interpreted the text to read either ‘unemployed’ or ‘inactive’. In the analysis of ALMPs, this distinction is important. Translation of these documents, which has been approved by SF, would remove these problems and would increase the efficiency of research, as it would remove the need for individual non-native researchers to separately interpret the same metadata for every new research project.

No information is provided on the robustness of the variables themselves

There are no contextual data given on how suitable the variables are for use in analysis. This then has implications for which variables researchers use in analysis and how they use them. For instance, variables often have missing values, but there is nothing to guide the researcher into the reasons underlying the lack of a value. In some cases, missing data may be interpreted as 0, however in other cases a missing value might indicate something other than a 0, for example that the individual in question belongs to a different population for whom it is not possible to obtain a value for them. The choice confronting the researcher is whether to exclude that individual or include them and impute a value (0 or otherwise). In some cases, the proportion of missing values is so high, that unless specific reasons for the underlying reason is known, it renders any information provided in that variable unusable. For example, in the TEM LMT data, a value which describes the occupation of the placement is missing in 84% of the records, with no guidance on why it should or not should be missing in the data. Similarly, there are variables which exist that seem to provide very similar information, but without guidance on where one should expect them to be similar and where differences should exist. This makes it possible for researchers to use different variables in analysis, which may then provide different answers, dependent on the degree of divergence between these variables.

1.4. Data comparability

This section reviews comparability of the data on the SF server to externally published information on labour market services, and also reviews the internal consistency of SF secure data that are available from different source datasets. In theory data should align between all the sources, so that a consistent and coherent story is told with whatever data are used. However, the process of deriving data from different sources and processing raw data means that often this is not the case. Any inconsistencies are compounded via a lack of documentation, both in the underlying raw administrative data but also in any published statistics.

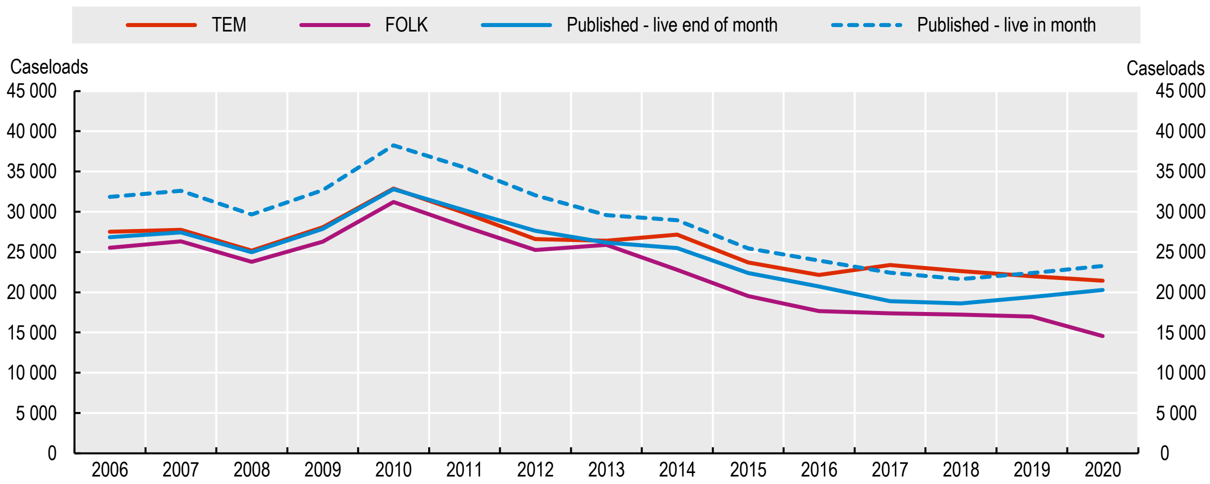
For example, whilst the free-of-charge public use statistical summary data available on the SF generally give good descriptions on what the data are, they do not provide sufficient detail on exactly how these data are calculated. To take the dataset “12u2 - Labour market training by province at the end of and during the month” as an example, the only information about its source is that it comes from the “Ministry of Economic Affairs and Employment, Employment Service Statistics”. No information is given on the underlying dataset or system that the statistics are produced from, or what the variable and/or processing rules for compilation are. Therefore, it is very difficult to determine with any certainty why numbers might be different from those contained in the dataset.

1.4.1. Administrative data on SF do not match published data

It is not straightforward when using the SF research data to be certain that the data used accurately reflect the administrative caseload for the item in question. Figure 1.1 provides an example of this looking at participants in LMT. Producing caseloads using the TEM data provide a good match to the published statistics until around 2013, whereupon a gap between the two series opens up, averaging around 2 500 participants per year. The same caseloads produced using the FOLK LMT data are around 1 500 participants lower per year until 2013 and then widening to around 3 000 after this point. In neither case it is obvious as to why this might be. Both sets of data on SF are spells data, which record training start and end dates. There are no records which "update" spells, to augment either start or end dates, so there are no obvious adjustments that can be made to refine data to make them closer to the published statistics. Neither is it obvious why the FOLK data series should be lower than that of the TEM data, when the FOLK data should be derived directly from the TEM data.

Figure 1.1. LMT caseloads do not precisely match published statistics

LMT caseload estimates using TEM and FOLK datasets against published estimates



Note: TEM and FOLK calculated as 'live at end of month' so should be most similar to the solid blue line. The published 'live in month' is provided for reference.
Source: TEM and FOLK are OECD calculations using Statistics Finland datasets. Published figures using SF free-of-charge public dataset 12u2, update 20220621 08:00

1.4.2. TEM SMT data contain information that is not used for published caseload statistics

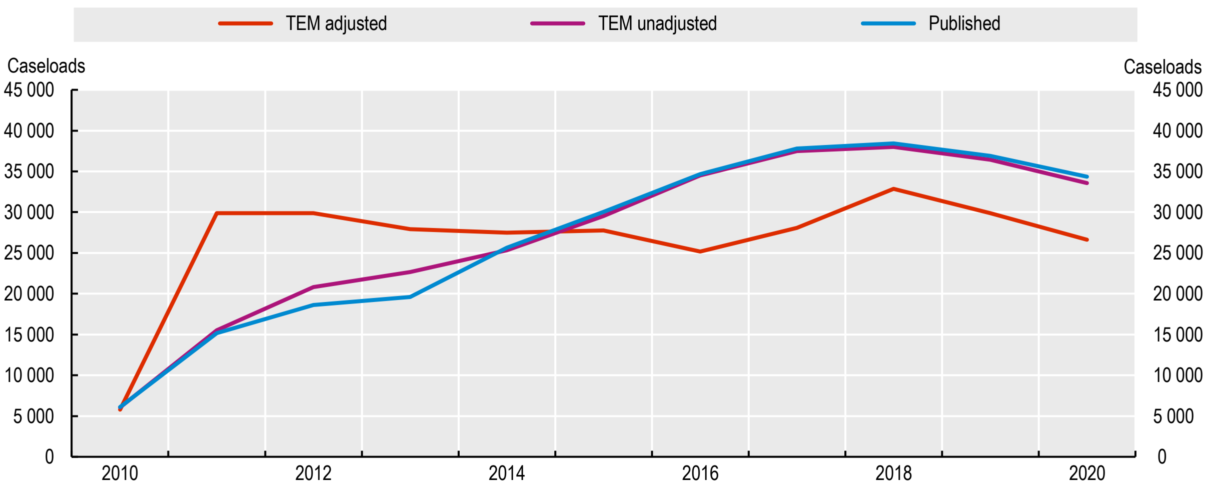
Building on the previous example, there are also examples where additional data exist and their incorporation into any ALMP participation figures is not straightforward.

The TEM SMT data available on SF are part of a larger suite of data that records activities on a jobseekers file. There are different entries for SMT. Broadly these relate to "starting", "pausing", "ending" and "ineligible" for SMT. Depending on how these records are analysed provides different estimates of how many people are participating in SMT. Figure 1.2 demonstrates this issue. The published statistics match almost exactly to caseload estimates that utilise the TEM SMT data and make no adjustments relative to records which may provide more information on a particular spell. The "TEM unadjusted" series simply takes the start and end dates for all records which are denoted as a "start" for SMT. This does not account for either explicit records which specify an "end" for SMT and provide a new end date, neither does it make any adjustment for records which are chronologically later and provide either updated start or end dates (these records may also be recorded as SMT “starts”). In this way it represents an upper bound on participation. This is because spell data are considered individually and the only adjustment is that where an individual has potentially two (or more) spells overlapping in the same period, the individual is only counted once in that period. The “TEM adjusted” series provides an estimate of the SMT caseload when rules on overlapping or conflicting spell data are introduced. For example, if a record for an “SMT end” is found, the end date of the previous SMT spell is updated with this new date. The difference in the estimate caseloads is around 5-10 000 across the time period, since the introduction of SMT.

There is no guidance in the meta data as to which of these series presents a more accurate picture of the number of participants on SMT. The fact that the published statistics match closely to the unadjusted series would suggest that it is this series which is more accurate. However, the question must then be asked of what precisely these additional data should be interpreted as if anything at all. Better guidance on use and suitability would be beneficial to ensure that researchers can be confident that they are analysing the correct data and that different sets of researchers use the same data when trying to answer questions on SMT.

Figure 1.2. Published SMT data do not make adjustments based on updated records

SMT caseloads using TEM datasets on SF server compared to published data



Note: TEM unadjusted produces monthly caseload data using spell start and end dates from TEM. TEM adjusted redefines spell end dates using either explicit records which denote ‘end’ of an SMT spell, or by using later records with the same start date and different end dates.
Source: TEM adjusted and unadjusted FOLK are OECD calculations using Statistics Finland datasets. Published figures using SF free-of-charge public dataset 12u9, update 20220621 08:00

1.4.3. TEM and FOLK SMT records do not match

To continue on from the previous comparisons, there are mismatches between the information available on the TEM and FOLK datasets on training. In general TEM identifies more claimants than the FOLK datasets, but this is not universal (e.g. some records can also be identified on FOLK but not TEM). Table 1.2 demonstrates this issue for SMT, where TEM records 41.5% more records than its FOLK counterpart.

Further analysis was conducted to attempt to account for the differences. To describe this analysis, first the identification methods for each dataset need to be described. The FOLK dataset analysed was the unemployment spells dataset. This records a reason for exiting unemployment as “voluntary studies on unemployment benefit”. This is the decode for an SMT start. The TEM dataset “statements of a person’s employment policy" (short dataset name "ytvpoll") records SMT starts in the variable "lausno" and also provides the expected end date.

Table 1.2. TEM data identifies more SMT than FOLK data do

SMT spells by data source

	TEM only	FOLK only	Both	Total
Number of SMT records	54 015	3 222	119 100	176 337
Row Percentage	30.63	1.83	67.54	100

Note: All training records identified in the data. Years restricted to 2012-2018 to ensure same time period coverage available on both datasets. FOLK data were only changed in 2012 to identify exits to SMT.
Source: OECD analysis of Statistics Finland TEM and FOLK data on self-motivated training spells.

Generally, the date of exit from the FOLK dataset (i.e. The last day of unemployment) is precisely one day prior to the SMT "start date" in the TEM data. "TEM only" SMT cases were first matched exactly on date-1 dates and then a further sense check was conducted by matching the TEM case to the closest SMT case found.

The analysis conducted to look at the cases only identified on TEM (54 015 cases) revealed:

- 17 029 - Matched to a different FOLK reason for exit (matched to exactly one day prior to LMT start). Of which unemployment end code of "other training" – 14 265. This looks like it possibly relates to incorrect FOLK coding for unemployment exits.
- 200 - No FOLK basic record in the same year (proxy for not resident in Finland at year end, which defines sample of the FOLK unemployment data)
- 36 786 - No reason found

Fully 68% of the TEM records remain unaccounted for and the reason why they do not appear on the FOLK data is unknown. It is possible that SMT start dates do not occur at the end of a period of unemployment. To check whether FOLK records which relate to "job search" rather than unemployment could explain some of this difference, the same principle of matching to the closest spell was conducted, however only 1% of remaining cases were matched to a job search spell ending one day earlier. This means that for those TEM SMT cases that remain unmatched, they start neither near the end of an unemployment spell nor the end of a job search spell.

This mismatch is not a problem per se, but it relies on making a choice between two datasets and having knowledge of these differences in the first place. SF charges researchers for each individual dataset accessed. It is feasible that a researcher chooses only to use FOLK datasets and the meta data for these FOLK datasets does not mention the incongruence with the TEM dataset on SMT. Therefore, a researcher could continue in good faith identifying only those cases available on FOLK, which may not be representative of the broader population of SMT participants.

1.4.4. TEM and FOLK LMT records do not match either

A similar exercise was conducted on the LMT records, with even less success. In fact, here the analysis is simpler, as both TEM and FOLK datasets contain LMT spell data- they both have start and end dates for the training itself.

To again try to determine why TEM cases could not be matched to FOLK cases, additional analysis was undertaken. To check the population consistency point, TEM only cases were matched to FOLK basic data in the same year. This yielded 4 512 cases which could not be matched to a FOLK basic record, suggesting that a lack of residency may have been responsible (so that the individual would not have met the criteria for inclusion in the FOLK LMT dataset).

Table 1.3. TEM also identifies more LMT participants than FOLK does

LMT spells by data source

	TEM only	FOLK only	Both	Total
Number of LMT records	340 482	27 419	779 256	1 147 157
%	29.68	2.39	67.93	100.00

Note: All training records identified in the data. Both means records are recorded on TEM and FOLK datasets. Data compared across the years 2006-2017
Source: OECD analysis of Statistics Finland TEM and FOLK data on self-motivated training spells.

There was no other obvious mechanism by which cases would not occur in both datasets. In the TEM data, there are numerous very short duration LMT courses, of one day in length. Indeed, of the 340 482 TEM LMT spells not matched to FOLK, 129 106 of them had a duration of one day. However, whilst this represents almost 40% of the unmatched cases, it doesn't account for the other 60% and there are also LMT courses of one day in length in the FOLK LMT records. So it offers nothing in conclusive proof in the discrepancy between the two data sources.

In order to provide researchers with better guidance on which of these data sources to use (and in which circumstances) some guide of addition to the metadata documentation would be necessary to detail the reasons underlying these differences.

2 Data linking and preparation

2.1. Data linking process

The data used to conduct the impact evaluation of SMT and LMT come from several sources (for an overview, see chapter 1), this section sets out how these data were linked and what cleaning and preparation was made to conduct the analysis.

The individual level datasets held by SF are all linkable using a unique individual identifier. This variable is an encrypted version of an individual's social security number and ensures that individuals can be easily linked, in a secure manner, which precludes the ability for this linking variable to identify the individual in any way. The existence of this variable on all of the individual level datasets mean it is a simple task to join different information about an individual together.

The possibility to seamlessly link datasets together provides a rich understanding of individuals' participation in self-motivated and labour market training, their socio-demographic characteristics, their past unemployment history and their labour market outcomes (including wages and occupations).

Table 2.1. Data retained and linked for the main analysis

Data needs	Source	Periodicity	Structure
Jobseekers	FOLK- basic dataset (*tua_perus; variable: ptoim1*)	Yearly: Identifies if individuals are unemployed at the last week of each year	Panel
Unemployment spells (dates) / History in unemployment	FOLK - unemployment spells (tyottomat)	Every time there is an event (= unemployment)	Spells
Participation In SMT	TEM – labour market policy statements (*ytvpoll*)	Every time there is an event	Spells
Participation in LMT	TEM – labour market training spells (*ytyokou*)	Every time there is an event (= training)	Spells
Outcomes: employment status, earnings and occupation	FOLK- data on employment spells (jua_tkt21)	Yearly: earned income and occupation available for the ongoing employment spell at the end of each year	Spells (ongoing at the end of the year)
Controls	FOLK - basic dataset (*tua_perus*)	Yearly: Individual's characteristics at the last week of each year	Panel

Note: The table provides an overview of the data used in the main analysis in (OECD, 2023[1])

2.1.1. Linking can also improve qualitative insights

This data linking extends to provides richer qualitative insights into training provision, alongside the broader counterfactual impact analysis. Linking of TEM administrative data on SMT to education data on student enrolments and FOLK basic data allows a more detailed breakdown on how individual characteristics vary by the type of course studied. Annex Table 3.A.1 shows that vocational and further SMT courses tend to be undertaken by students who live in rural municipalities (relative to either Bachelors or Masters level courses). Vocational and further course participants are also less likely to already have a tertiary level qualification. Those undertaking vocational upper secondary level training are also less likely to be Finnish or to speak Finnish, suggesting that these courses are utilised more by immigrants to Finland.

These data also go someway to illustrating what data are missing in the analysis of LMT. Without recourse to knowing any of the details on what type of course individuals are taking, it is not possible to analyse whether or how different types of individuals select different training courses.

2.2. Sample selection

The database used for the evaluations by the OECD (2023[1]) was constructed by imposing some restrictions and restructuring some of the data:

- Only individuals that can be observed in the data up to four years are kept. This implies first that individuals that disappear from the FOLK data are not considered. In fact, from individuals observed in a given year, around 5% are no longer in the data four years later (this is mainly due to mortality and migration). As the outcomes studied are only observable until 2018 this also implies that only individuals observed up to 2014 can be kept in the sample.
- Similarly, since controlling for past employment/unemployment history is crucial for the empirical strategy (see section 3.2.2), only individuals for whom outcomes can be traced over the past two years are part of the sample. Because SMT started in 2010, therefore only individuals observed as of 2012 are kept in the data.
- Only unemployed persons are kept as they are the population of interest. More precisely, due to the yearly-panel structure of the FOLK data, only individuals that have an ongoing unemployment spell at the end of years 2012-14 are considered for the analysis.
- Since one of the value-added of the evaluation conducted by the OECD is to explore occupational mobility, it is necessary to observe individuals' occupation at the starting observation point (before the start of the programmes). For the unemployed this occupation is recovered as the occupation of their last employment known in the data. Therefore, individuals with missing previous occupation are excluded from the analysis.
- The data were reorganized such that the unit of observation is an individual unemployed person. In fact, FOLK basic data are structured as a yearly panel presenting one observation per individual and per year, thus, for instance, an unemployment spells that lasts over two consecutive years would appear in two observations. The data were thus re-structured to keep only one line per individual to avoid duplicating information and to ease the implementation of the empirical strategy. For individuals having more than one unemployment spell during 2012-14 only the last spell was kept (information on the previous unemployment spell appears in the previous unemployment history variables).

The final sample for the evaluation consists of around 370 000 jobseekers with an ongoing unemployment spell the last week of years 2012-14. Among them, more than 13 000 entered SMT and more than 16 000 LMT (of more than three months) in the following year.

The restrictions made and the data cleaning process have consequences in terms of the representativeness of the sample used for the analysis. First, the sample consists in individuals we observe in unemployment at the end of each year, the analysis thus excludes jobseekers that have episodes that do not overlap with the end of a calendar year. This first exclusion yields to a sample of individuals that has a higher recurrence to unemployment in terms of both days and spells, men are over-represented and the population is slightly older on average than the overall unemployed population (Column 2 of Table 2.2). Second, to study occupational mobility, only jobseekers who have a known previous occupation are kept in the sample. Jobseekers that have previous occupation are jobseekers that have past work experience, and thus they have on average less spells and days spent in unemployment in the past, they are younger and less likely to be woman (Column 3 of Table 2.2). Furthermore, keeping jobseekers with previous occupation, allows to get more information on the level and field of education, as the category "unknown" is smaller within this sub-group. Last, jobseekers without missing values in the outcomes studied also differ slightly from other jobseekers (Column 4 of Table 2.2).

Table 2.2. Jobseeker's characteristics vary with sample restrictions

Jobseekers' characteristics	Mean - all unemployed	Mean - unemployed at the end of the year	Mean - unemployed at the end of the year with previous occupation	Sample mean
History in unemployment				
Unemployment duration until 31 of December (in current spell)	274.843	284.086	268.122	271.752
Number of unemployment spells in previous year	1.142	1.217	1.186	1.201
Number of days in unemployment in previous year	76.895	94.793	88.389	89.321
Number of unemployment spells in previous 2 years	1.852	1.999	1.983	2.01
Number of days in unemployment in previous 2 years	150.536	187.48	176.924	178.801
Demographic characteristics				
Finnish national	0.94	0.937	0.962	0.965
Woman	0.462	0.432	0.424	0.425
Age	38.582	40.639	42.692	42.931
Number of children	0.752	0.709	0.694	0.694
Car ownership	0.486	0.492	0.566	0.572
Marital status				
Unmarried	0.525	0.501	0.458	0.455
Married	0.339	0.344	0.374	0.377
Divorced	0.126	0.143	0.156	0.156
Widowed	0.01	0.012	0.012	0.012
Status in the family				
Not belonging to a family	0.312	0.329	0.314	0.311
Head of the family	0.205	0.216	0.245	0.248
Spouse	0.166	0.16	0.163	0.165
Child	0.102	0.093	0.061	0.06
Head of cohabiting family	0.1	0.099	0.112	0.112
Spouse of cohabiting family	0.097	0.083	0.088	0.087
Unknown	0.017	0.019	0.017	0.016
Type of housing				
Detached house	0.374	0.373	0.413	0.417
Terraced house	0.127	0.126	0.133	0.134
Block of flats	0.461	0.461	0.418	0.414
Other building	0.02	0.02	0.019	0.018
Living alone	0.261	0.281	0.272	0.271
Municipality type				
Urban	0.723	0.721	0.703	0.699
Semi urban	0.146	0.146	0.155	0.157
Rural	0.131	0.133	0.143	0.145
Language				

Jobseekers' characteristics	Mean - all unemployed	Mean - unemployed at the end of the year	Mean - unemployed at the end of the year with previous occupation	Sample mean
Finnish	0.86	0.858	0.898	0.902
Swedish	0.028	0.025	0.028	0.027
Other	0.112	0.117	0.074	0.071
Level of education				
Upper secondary or less	0.531	0.527	0.539	0.546
Post-secondary non tertiary education	0.005	0.004	0.005	0.005
Short cycle tertiary education	0.066	0.07	0.084	0.084
Bachelors or equivalent	0.094	0.085	0.098	0.096
Masters or equivalent	0.071	0.064	0.073	0.07
Doctoral or equivalent	0.005	0.005	0.006	0.005
Unknown	0.229	0.245	0.195	0.193
Field of education				
Generic	0.08	0.061	0.048	0.046
Education	0.013	0.009	0.01	0.01
Arts and humanities	0.056	0.056	0.057	0.055
Social sciences	0.012	0.011	0.011	0.011
Business	0.106	0.108	0.122	0.12
Natural sciences	0.01	0.009	0.01	0.01
ICT	0.032	0.033	0.033	0.033
Engineering	0.253	0.274	0.3	0.305
Agriculture	0.027	0.029	0.03	0.03
Health and welfare	0.073	0.057	0.068	0.07
Services	0.101	0.1	0.108	0.11
Unknown	0.237	0.252	0.201	0.199
Profession of previous occupation				
Armed forces	0.004	0.005	0.001	0.001
Managers	0.017	0.02	0.015	0.015
Professionals	0.117	0.103	0.109	0.107
Clerical support	0.069	0.072	0.063	0.063
Service and sales	0.216	0.193	0.22	0.219
Skilled agricultural, forestry and fishery	0.024	0.029	0.017	0.017
Craft and related trades	0.183	0.199	0.21	0.213
Operators and assemblers	0.111	0.121	0.116	0.118
Elementary occupations	0.133	0.143	0.126	0.127
Observations	914 207	594 581	409 672	370 882

Note: This table concerns the 2012-14 period. The first column of the table shows average jobseeker's characteristics for all individuals that had an unemployment spell over 2012-14. The second column presents the mean characteristics of individuals with an ongoing unemployment spell the 31st of December of years 2012-14. The third column presents the characteristics of individuals with an ongoing unemployment spell the 31st of December of years 2012-14 that have a previous occupation known in the data. Finally, the fourth column presents the characteristics of jobseekers in the final sample, and thus the ones that, among jobseekers in the third column, do not present missing values for any of the outcomes studied.
Source: OECD calculations based on FOLK datasets.

3 Methodology and robustness of the causal impact evaluation of training

This chapter outlines a number of important features of the analysis and methodology used in the OECD analysis (2023[1]) of labour market training (LMT) and self-motivated training (SMT). It details a schematic overview of the process of framing policy analysis in terms of its objectives and how these relate to inputs, activities, outputs and outcomes. This motivates a discussion of the outcomes that are possible to analyse in the impact evaluation of LMT and SMT and describes how detailed information on occupations has been compiled to provide a tractable insight into how training influences participants' occupational choices. To provide a robust impact of the causal effects of both programmes, analytical techniques that overcome potential differences between participants and non-participants are essential. The techniques used in the assessment of LMT and SMT are detailed, alongside the underlying assumptions they make which need to hold true for the results to provide a true causal impact of the programmes. Several robustness checks are then detailed, to provide insight into how sensitive the results are to different methods of calculation and whether the underlying analytical assumptions bear scrutiny.

3.1. The impact evaluation is set up in a well-defined results chain framework and leverages a rich set of labour market outcomes

This section describes how constructing a conceptual framework for policy analysis helps to provide structure into thinking about how the objectives for policy are expected to work through a chain of interconnected processes, which culminate in a set of measurable outcomes to evaluate success. The outcomes that are defined and measured for the LMT and SMT analysis are then described, alongside a more detailed look into how occupational data are used to bring insight into the dynamics of occupational change following training participation.

3.1.1. Results chain framework

Impact evaluations are essential to develop evidence-based policy as they inform on the effectiveness of programmes in achieving specific targets. In order to specify which questions an evaluation intends to answer it is important to understand first the causal logic behind the programmes; from identifying the resources available to defining the results to be accomplished. More precisely, this means determining what are the inputs, the activities, the outputs and the outcomes of the programmes (Gertler et al., 2016[6]). In the context of this study, the "results chain" framework can be defined as follows:

- Inputs: the resources available to LMT and SMT programmes. Funds used for training programmes from The Ministry of Economic Affairs and Employment (TEM) as well as from The Ministry of Education and Culture (OKM) regarding vocational training programmes (that are part of LMT) and higher education provided to jobseekers under SMT[2]. These also include TE Offices staff and resources to administer the programmes.

- Activities: activities that convert inputs into outputs of the programmes. These include the training programmes (in the case of LMT) and the formal education courses (in the case of SMT).
- Outputs: tangible goods and services generated by programme activities. These include the number of individuals who completed training (for LMT) or who obtained the target diploma (for SMT). Outputs are monitored by TEM and OKM.
- Net outcomes (impacts): the effects that the programme achieves after the target population has received or been exposed to the programmes' outputs and activities, after taking into account counterfactual outcomes of participants had they not participated. These effects can be measured along different dimensions; the ones examined in this study are detailed in next sections.

Inputs, activities, and outputs constitute the implementation phase that is usually under the responsibility of the implementing bodies (TEM and OKM). CIEs thus focus on determining net outcomes only. The main question they intend to answer is whether there are changes that can be causally attributable to the programmes. By measuring this causal impact, it is possible to assess if the objectives of the programmes were met. The next sections describe the outcomes on which the impact of LMT and SMT is measured.

3.1.2. Definition of outcomes

In the case of Finland, most outcomes of interest can be tracked trough the rich and comprehensive FOLK datasets. Since FOLK datasets are built as snapshots of individuals at the last week of every year, outcomes are tracked at this calendar point over up to four years from the starting observation point. The following outcomes are examined:

- Probability of being employed. This probability is measured using a binary outcome variable, which is equal to one if the individual is employed at a given point in time and equal to zero otherwise. This binary outcome comes from the variable that indicates the employment relationship during the last week of the year. It classifies the population according to whether they are employed, unemployed, aged below 14, students, pensioners, conscripts or in another situation outside the labour force.
- Months of employment in year. Calculated as an integer which is categorised into months using the underlying number of days worked.
- Annual earned income. Yearly earned income comes from tax data and is available for the study until 2018.
- Monthly earnings. It is measured as the yearly-earned income divided by the number of months in employment. Yearly earned income comes from tax data and is available for the study until 2018.
- Probability of changing occupation. This probability is measured using a binary outcome variable, which is equal to one if the individual's occupation at a given point in time is different than at the starting observation point and equal to zero if it is the same. Occupational codes in FOLK data are given at the 3-digit level following the International Standard Classification of Occupations (ISCO). Individual's occupational code is primarily formed by the occupational title of employment relationship during the last week of the year. In addition, data on industry and sector of the employer and education data can be used for determining occupation. In this research, for unemployed

[2] Note that the way to take into account the resources available for SMT, and particularly the cost of providing education, is not straightforward. On one hand one could consider that there are no additional marginal costs for SMT provision for unemployed jobseekers as the educational programmes would take place anyways. On the other hand, when comparing the costs (and benefits) of different training measures for the unemployed, it seems crucial to take somehow the costs of providing the training into account to make the comparison more accurate.

individuals the occupation is defined as the one of their last employment known in the data. This occupational classification is used since 2010.

- Occupational mobility. The analysis maps the occupation of individuals onto an occupational index, which is interpreted as a "job ladder". It is measured in monetary units and in percentile rank. The construction of the index is detailed in the next section.

3.1.3. The occupational index was constructed based on the average earned income over the 2012-18 period

In addition to evaluating the effect of training programmes on typical labour market outcomes such as employment and earned income, the OECD has also analysed occupational mobility. The OECD (2023[1]) analysis investigates whether jobseekers end up in better quality occupations as a result of participation in training programmes (i.e. moving up the "job ladder"). To measure the quality of occupations an occupation index is calculated from observed earned income following the approach adopted by (Laporšek et al., 2021[7]). The next steps are followed:

1. Compute the monthly earnings for all employed individuals over the 2012-18 period by dividing the earned income in a calendar year over the months of employment.
2. Average out the monthly earnings at the level of 3-digit occupations. This gives the occupational index expressed in monetary units.
3. Attribute a percentile rank to the occupations according to their average monthly earnings to obtain an occupational index in percentiles (See the relationship between the occupational index in income and in percentiles in Figure 1 of Annex 3.A).
4. Allocate this occupational index to all employed and unemployed individuals matching on their occupations. For unemployed people, the occupation assigned is the one of the last employment observed in the data. This implies that unemployed persons without any work experience are excluded from the sample.

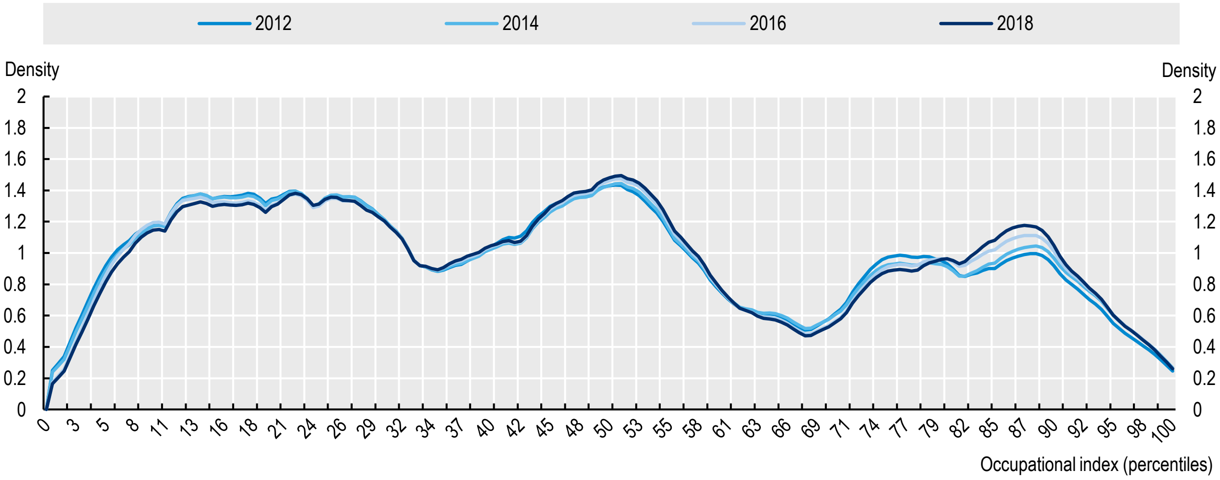
Note that some adjustments were made before obtaining the final occupational index. Outliers in terms of earnings were dropped at the individual level (top and bottom 1%). Furthermore, only occupations for which the average earnings are properly estimated were kept: occupations with less than 700 observations over the 2012-2018 period were dropped (four occupations[3]). As a result, the occupational index ranks 122 occupations.

The distribution of the occupational index among employed in Finland is constant over the 2012-2018 period (see Figure 3.1). Individuals are distributed all over the spectrum of occupations, with some preferred occupations. They are more concentrated in the bottom/middle occupations and less concentrated at the top. As compared to employed individuals, unemployed individuals are disproportionally represented in low-occupations (see Figure 3.2). On average, unemployed individuals have an occupational index that is approximately 7.7 percentage points lower than employed ones, corresponding with a drop of EUR 272 on the average occupational monthly earnings.

[3] The occupations dropped are likely to reflect inaccuracies in the occupational codes. Two of them are coded with only 2-digits.

Figure 3.1. The occupational index in Finland is constant over the 2012-18 period

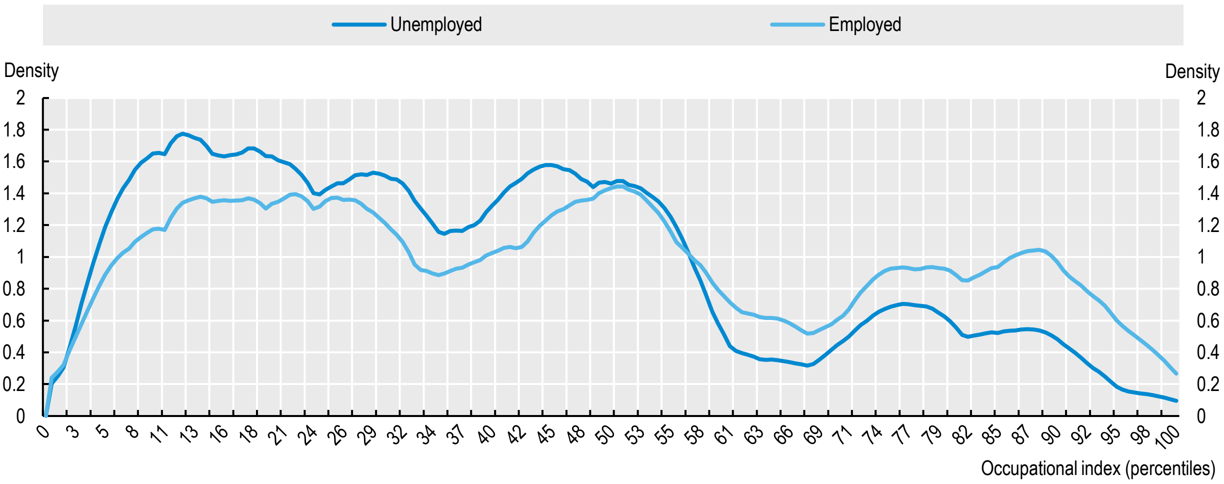
Evolution of the occupational index for employed individuals in Finland, 2012-2018



Note: The heights of the lines indicate the relative share of individuals in occupations whose index in percentiles are on the horizontal axis. The distributions are calculated for all individuals who were employed the 31st of December of each calendar year. Observations with index value above 100 are excluded from the kernel density chart.
Source: OECD calculations based on FOLK datasets.

Figure 3.2. Unemployed Individuals are overrepresented in lower-paid occupations and underrepresented in higher-paid occupations

Occupational index distribution for employed and unemployed individuals in Finland in December 2014



Note: The heights of the lines indicate the relative share of individuals in occupations whose occupational index in percentiles are on the horizontal axis. The distributions are calculated for all individuals who were employed or unemployed at the end of 2014. Observations with index value above 100 are excluded from the kernel density chart. For unemployed individuals the occupation is the one of their last employment known in the data.
Source: OECD calculations based on FOLK datasets.

3.2. Description of econometric techniques used and rationale for their use

OECD (2023[1]) uses the potential outcomes framework, sometimes termed the Neyman-Rubin Causal framework (Neyman, 1923[8]; Rubin, 2006[9]), which expresses the causal inference problem as having to estimate the impact of a programme never having observed the alternative state in which the individual did not participate in the programme. This framework has a widespread application in empirical analysis of labour market programmes (for example see Card, Kluve and Weber (2018[10]) for a summary of papers using such a framework). This is implemented primarily using a technique called propensity score matching, which is discussed in the following section. For the analysis which looks at how the introduction of SMT financing for longer-term education changed jobseekers' training choices, a difference-in-difference approach is used in conjunction with this technique to help to address potential unobserved differences between groups.

For these techniques to be valid, they rely on a set of underpinning assumptions. Because the alternative state of an individual (their "counterfactual") is never observed, the techniques rely on comparisons to other groups of individuals that serve as appropriate proxies for these never observed states. The different techniques make different assumptions on how they assure these proxies represent the true counterfactual. These assumptions are discussed in the following sections.

3.2.1. Counterfactual impact assessments are needed to compare similar individuals

CIEs of ALMPs pose many methodological challenges. One of the main ones is the selection bias issue. To evaluate a given ALMP, it is not sufficient to compare the outcomes of its beneficiaries with those of non-beneficiaries directly. Indeed, beneficiaries are often "selected", i.e. they do not have the same characteristics as non-beneficiaries and are therefore not comparable (e.g., beneficiaries might be better informed or more motivated). In order to isolate the causal effect of a policy, it would be ideal to be able to compare the outcomes of the beneficiaries of this programme with the outcome that these same individuals would have obtained if they had not benefited from this programme. However, it is not possible to observe what would have happened to the beneficiaries in the absence of the policy.

The objective of CIEs is therefore to compare the outcomes of individuals who have benefited from a programme (the treatment group) with those of a set of individuals as similar as possible (the control group). The only difference between the treatment group and the control group is that the latter did not participate in the programme. The control group therefore provides information on "what would have happened to the individuals subjected to the intervention if they had not been exposed to it", that is the counterfactual case.

Two approaches can be distinguished among CIEs: experimental evaluations, also called randomised controlled trials (RCTs), and quasi-experimental evaluations. In an RCT, the treatment and control groups are selected randomly from a given population. If the selection process is truly random, the characteristics of the individuals in the two groups do not differ on average: the groups are therefore statistically equivalent. Comparing the outcomes of these two groups at the end of the programme allows the causal effect of the programme to be isolated. However, even if RCTs are an ideal framework for evaluating public policy, they are not always feasible. RCTs require substantial financial and logistical resources, and programmes are often designed without following an experimental protocol. In addition, it can be undesirable to limit access to a potentially beneficial ALMP to a limited part of the population for the sole purpose of evaluating it.

SMT and LMT programmes in Finland do not follow an experimental protocol. More precisely, either PES staff play an important role in directing individuals to training programmes, either individuals self-select (mostly in the case of SMT) , therefore those who could probably benefit the most from the programme or those who are anyways closest to the labour market could be selected. Participation is thus not randomised and beneficiaries differ considerably from the rest of the unemployed population (see (OECD, 2023[1]) for more details).However, there is still a need to evaluate such programmes and, to this end, quasi-experimental methods can be considered. Quasi-experimental approaches essentially attempt to mimic the randomisation process described above by constructing a control group as close as possible to the treatment group, so that they are statistically equivalent ex-ante.

Different quasi-experimental techniques can be deployed to conduct CIEs, the chosen technique will depend on the context of the programme and its implementation. For instance, in the case there is a well-defined eligibility criterion for participation a regression discontinuity design (RDD) methodology can be applied. The rationale is that participants and non-participants at either side but close to the eligibility threshold (for example, age or income) should be similar in all respects except their participation in the programme. Similarly, the difference in differences strategy states that even if participants and non-participants outcomes are not comparable in levels, it is possible that the evolution of the outcomes of non-participants over time follows the same trend than the one of participants in the absence of treatment. Therefore, the causal impact of a programme is retrieved by comparing the changes experienced by participants and non-participants.

Regarding SMT and LMT, no institutional rule affects individual's participation exogenously. To assess the impact of these training programmes separately the method chosen is nearest neighbour propensity score matching. This method builds a control group by matching one by one, treated individuals with non-treated individual based on observable characteristics. Furthermore, to study the interaction between the two types of training programmes a difference in differences strategy is used. It exploits the introduction of self-motivated training and compares the setting before, when only labour market training existed, and after, when both labour market and self-motivated training became available to jobseekers.

3.2.2. Propensity score matching is used to account for the differing composition of the treatment and control groups

In order to account for the differing composition of participant and non-participant jobseekers in SMT and LMT, an econometric approach that matches individuals on observable characteristics is adopted. This approach, called propensity score matching, attempts to ensure the comparability of the treatment and control groups and provide reliable estimates of the effects of both training programmes. Specifically, a rich set of observable characteristics is used to identify individuals with similar probabilities of enrolling into these training programmes. Individuals are then paired with similar individuals based on this probability and their outcomes are compared. Such an approach – based on a so-called propensity score – is commonly used in the literature to address the difficulty of otherwise accounting for a wide array of additional personal characteristics (Card, Kluve and Weber, 2018[11]).

The calculations of the propensity score takes into account three main groups of variables to ensure comparability:

- **Exogenous demographic characteristics:** age, gender, citizenship, level of education, language spoken, etc.
- **Past employment/unemployment history (lagged outcomes):** employment status the year before, number of spells and days spent in unemployment during the past year and the past two years, industry of the occupation the year before (1-digit ISCO code), etc.
- **Time controls:** dummy variables indicating the quarter of registration into unemployment.

Since due to the structure of FOLK datasets the main outcomes of interest only are observed at the end of every year, the sample used in the evaluation consists of individuals that have an ongoing unemployment spell at the end of a given year $t = 0$. More precisely, past unemployment history is controlled for up to two years and outcomes are followed for up to four years, the sample studied comprises individuals with ongoing unemployment spells at the end of years 2012-14[4]. Among them, jobseekers are considered as treated if they start the training programme of interest somewhere between $t = 0$ and $t = 1$. The control group comprises individuals that did not enter a training programme at that time and that are matched to treated individuals through propensity score matching. SMT and LMT treatment effects are estimated from $t = 1$ to $t = 4$. Denoting Y_{it}^{d} labour market outcomes (such as employment or occupational mobility) for an individual (i) where $d = 1$ under treatment and $d = 0$ otherwise, the average treatment effect on the treated ($D_i = 1$) can be written:

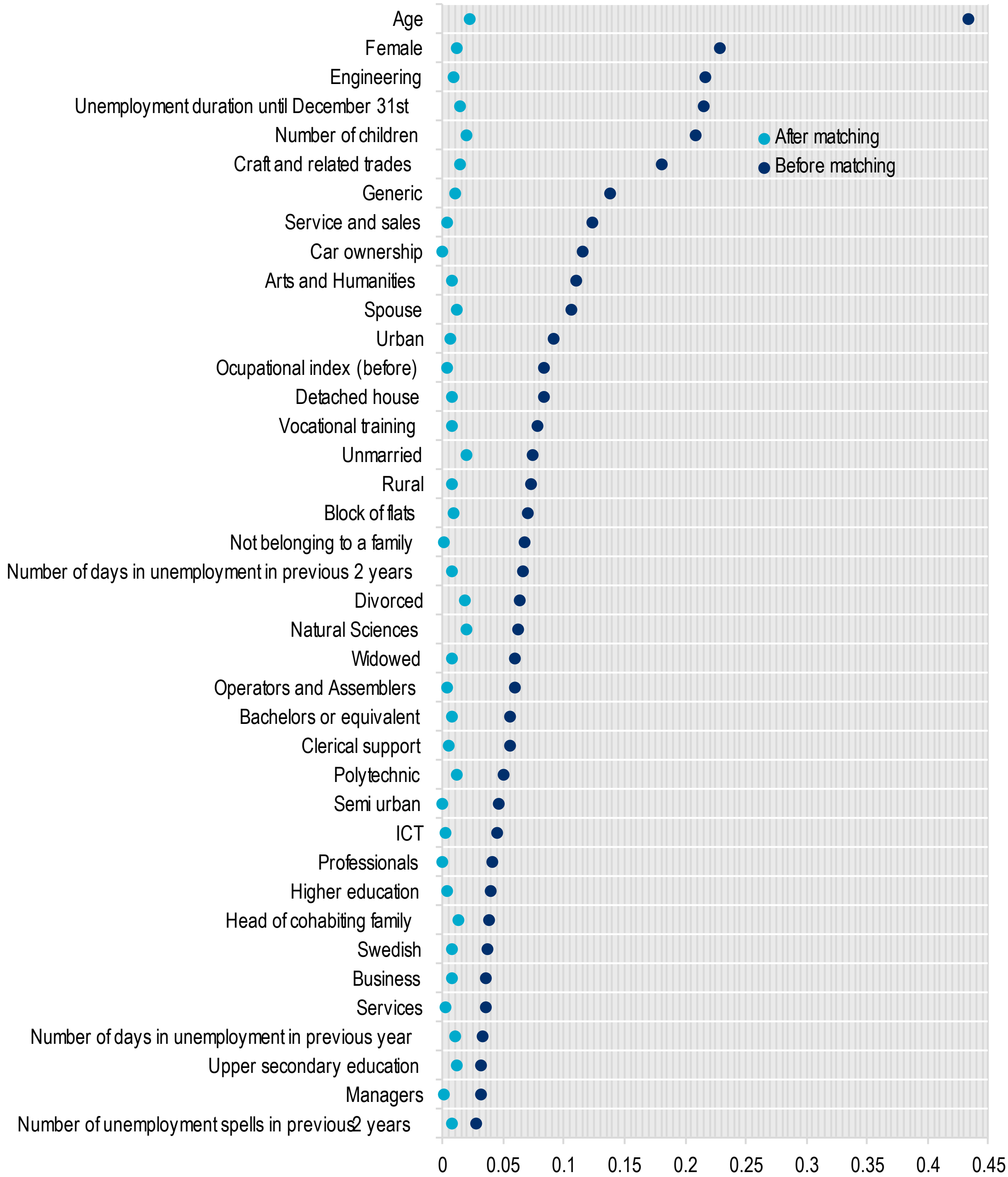
$$\gamma_t = E[Y_{it}^{1}|D_i = 1] - E[Y_{it}^{0}|D_i = 1]$$

The treatment effects obtained through propensity score matching can be interpreted as causal under the assumption that treatment and control groups are comparable conditional on observable characteristics. To test to what extent propensity score matching reduced differences between participants and non–participants, the standardised distance between treatment and controls is compared before and after matching across covariates (see Figure 3.3 and Figure 3.4). This balancing test shows that propensity score matching indeed improved the comparability of the treatment and control groups for both SMT and LMT. By construction the test can only be carried out on observable characteristics, thus unobservable confounding variables may still be unbalanced potentially leading to biased results. However, the extensive set of covariates used in this study include exhaustive information on jobseekers in almost all aspects that have been shown by the literature to be predictive of unemployment outcomes. In particular, labour market histories and earnings capture much of the information contained in usually unobserved variables (Heckman et al., 1998[12]; Caliendo, Mahlstedt and Mitnik, 2017[13]).

[4] Data on most outcomes is available for this study up to 2018.

Figure 3.3. Propensity score matching improved the comparability of the treatment and control groups for self-motivated training

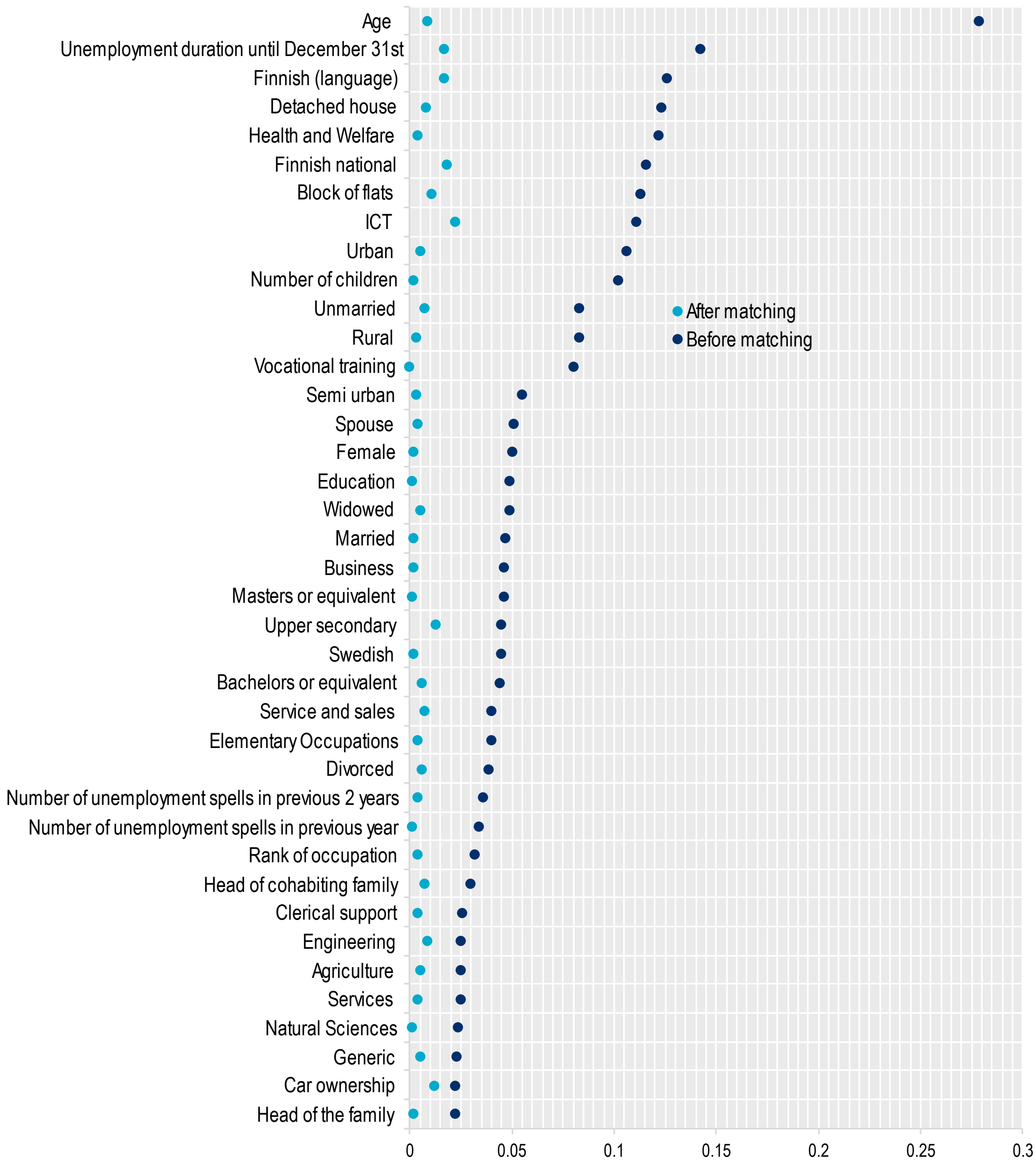
Standardised distance between treatment and control groups (0=no difference)



Note: Figure presents standardised differences for the variables used in the propensity score matching with the greatest standardised differences before matching. Standardised differences are calculated as the difference in means between the treatment and control groups for the matching variable divided by the square root of the sums of the variances for that variable. Unemployment duration until the 31st of December refers to the duration in the current spell. "Swedish" refers to being a native Swedish speaker.
Source: OECD calculations based on FOLK and TEM data.

Figure 3.4. Propensity score matching improved the comparability of the treatment and control groups for labour market training

Standardised distance between treatment and control groups (0=no difference)



Note: Figure presents standardised differences for the variables used in the propensity score matching with the greatest standardised differences before matching. Standardised differences are calculated as the difference in means between the treatment and control groups for the matching variable divided by the square root of the sums of the variances for that variable. Unemployment duration until the 31st of December refers to the duration in the current spell. Swedish refers to being a native Swedish speaker.
Source: OECD calculations based on FOLK and TEM data.

3.2.3. A difference in difference strategy is used to analyse the interactions between self-motivated training and labour market training

In the case that selection into programme is governed not only by observable but also unobservable individual characteristics that are correlated with the potential outcomes, then propensity score matching provides biased estimates of treatment effects. There is a greater potential for this to be the case when comparing individual outcomes across different years, where cohort and economic cycle effects may also be present. As an additional analytical fail safe against such issues, it is possible to combine propensity score matching with a difference-in-difference (DID) approach. The DID estimator means that only changes over time are compared, so that it removes all underlying time-invariant differences between groups. In this sense all (time-invariant) unobservable characteristics are controlled for, in addition to the observable ones controlled for by the propensity score matching. This is an approach that has previously been adopted by others (Heckman et al. (1998[14]), Smith and Todd (2005[15])).

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:

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

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

To investigate whether the introduction of education financed by SMT affected overall training outcomes, the first cohort of individuals that could participate in either SMT or LMT, was compared to the cohort of individuals in the year before, where only LMT was available. As for the separate analysis of SMT and LMT, the sample was defined as the pool of unemployed people at the end of the previous year. Therefore, the analysis constituted 2 groups of individuals:

- LMT cohort – individuals unemployed at the end of 2008 who participated in LMT in 2009
- SMT/LMT cohort – individuals unemployed at the end of 2009 who participated in SMT or LMT in 2010

Differences in the annual outcomes experienced by individuals at the same points in time, following a period of unemployment are then compared. If differences occur, then it can be said that the training set offered to one cohort changes labour market outcomes differently to the training set offered to the other cohort.

The following regression is estimated

$$\gamma_{it} = \alpha + \beta cohort_c + \delta period_t + \theta_t cohort_c * period_t + \varepsilon_{it}$$

Where

γ_{it} is outcome for individual I in period t

$cohort_c$ = 1 if unemployed in 2009, 0 if unemployed in 2008

$period_t$= integers between $t = -4$ and $t = 6$, defined as the difference between current year and start year in unemployment (i.e. equals 0 for 2009 cohort in year 2009 and 0 for 2008 cohort in year 2008). Enters regression as a series of dummy variables. $t = -4$ enters as the reference period, coefficient estimates for periods are all relative to this year (this serves as our measure for “differencing” the outcomes)

The coefficient of interest θc_t provides the estimate of the additional impact of $cohort_c$ in $period_t$. For the DID estimate to be valid, this should be insignificant in periods -3 to 0, so that the pre-treatment trends are similar between the control and treatment groups (this is known as the "parallel trends assumption").

Testing the validity of the parallel trends assumption

In addition to the simple DID estimator provided above, matching was used to provide construct a group of non-participants that can help to account for any potential violations of the parallel trends assumption. For example, consider a group of training participants that are older than the non-participants. Earnings tend to grow fastest when individuals are younger, before plateauing when individuals reach their 40s. If DID was used to compare these two sets of individuals, the fact that non-participant earnings grow more quickly may be related only to this age dynamic. The parallel trends assumption would not hold. By using matching, to ensure that participants were compared against non-participants with a similar age only, it can be used to restore the integrity of the parallel trends assumption. The robustness checks below present the impact of using matching in addition to DID alone. The results from conducting this analysis are reported in the next section.

3.3. Robustness checks

In addition to the main results presented in the OECD (2023[1]) impact evaluation, a number of robustness checks have been undertaken to determine whether the estimated results are sensitive to the particular econometric specification that has been used.

3.3.1. The main conclusions of the analysis remain robust to changing the data source

As explained in section **1.4**, the information contained in FOLK and TEM data presents some discrepancies. Importantly, the individuals identified as participating in labour market training or self-motivated training differ from one source to the other. This poses a problem while implementing the empirical strategy as individuals identified as belonging to the treatment group differ depending on the data source retained. Recall that the treatment group in the analysis consist of individuals having an ongoing spell at the end of years 2012-14 that enrol into one of the training programmes during the following year. Propensity score matching is then applied to match these individuals to similar individuals that did not enrol in the programmes. Using FOLK databases, the analysis would rely on 8 735 individuals identified as treated for self-motivated training and 13 691 individuals for labour market training (of more than three months) (see Table 3.1). Using TEM databases this numbers raise to 13 766 and 16 489 individuals respectively. Since the reasons for the discrepancies are not completely clear, it is important to make sure that the results obtained in the impact evaluation are not driven by the choice of the data source.

While the main results of this study use TEM data as the source to identify programme participation, Annex Figure 3.A.2 and Annex Figure 3.A.3 show how this results change when FOLK is the data source retained. Overall, the conclusions of the main analysis remain unchanged.

For SMT there is a lock-in effect followed by positive results in terms of employment and slight negative results for occupational mobility four years from the starting observation point. In the main analysis, the effect on monthly earnings remained slightly negative even at the last observation point and, when using FOLK data, this estimate becomes insignificant. However, both estimates are not statistically different. LMT exhibits positive effects on employment and no effects on occupational mobility. When using FOLK data the results on employment are slightly more positive and there is a positive effect on monthly income. However, as for SMT, the estimates measured with both data sources are not statistically different. Therefore, the results of the empirical analysis are robust to the use of a different data source.

Table 3.1.The number of individuals identified in the treatment group varies considerably depending on the data source.

Propensity score matching treatment group size according to the data source used for self-motivated training (SMT) and for (different durations of) labour market training (LMT).

	SMT	All LMT	LMT > 3 months	LMT > 6 months
FOLK	8 735	25 064	13 691	8 362
TEM	13 766	41 787	16 489	9 239

Note: This table summarizes the number of individuals identified as belonging to the treatment group when using nearest-neighbour propensity score matching according to the data source chosen.
Source: OECD calculations based on Statistics Finland's repository: FOLK and TEM datasets.

3.3.2. The main results are robust to changing the sample of interest

One of the main contributions of this study is to evaluate the effect of SMT and LMT on occupational mobility. However, to do so the sample used for the main analysis is restricted to individuals for whom a previous occupation is present in the data. As shown in Table 2.2, this restriction leads to a sample of individuals that has specific characteristics and thus is not representative of the universe of unemployed.

To understand the relevance of the main results found in this analysis it is thus important to look at how the results vary if individuals with no previous occupation are included in the sample of interest. Even if occupational mobility cannot be assessed for such a population, the effects on employment and earnings can still be measured.

Figures 3.A.4 and 3.A.5 of the Annex show that the main results on employment and earnings, for SMT and LMT respectively, remain almost unchanged when individuals with no previous occupation are included in the sample and are not statistically different than the ones of the main analysis.

3.3.3. The main results are robust to changing the estimator used

Many estimators can be used to compute the average treatment effect of the programmes through propensity score. Recall that propensity score measures the probability of being treated of individuals given their covariates. The inverse probability weighting estimator measures treatment effects by weighting the outcomes by the inverse of individual's propensity score. Another commonly used estimator is the kernel matching estimator; it matches each treated individual with a weighted average of all controls with weights that are inversely proportional to the distance between the propensity scores of treated and controls. Thus, it has the advantage of using all the information available, which leads to lower variance. However, a drawback is that observations used might be bad matches.

The literature has not identified a preferred estimator (Frölich, 2004[16]; Huber, Lechner and Wunsch, 2013[17]; Busso, DiNardo and McCrary, 2014[18]) mainly because the relative performance of estimators arguably depends strongly on features of the data-generating process which is unknown to the empirical researcher in practice.

This study employs a nearest neighbour direct matching estimator. It matches each treated individual with the nearest control in terms of propensity score. Given the large number of potential control individuals in the Finnish data (all the unemployed), this technique allows to get as close as possible to an ideal exact matching scenario in which each treated would be matched with a control having exactly the same propensity score. In fact, this estimator exhibits the smallest bias regardless of sample size (Huber, Lechner and Wunsch, 2013[17]). Furthermore, the large sample size also allows reducing the variance, whose potential large size is the main drawback of this estimator.

Annex Figure 3.A.6 and Annex Figure 3.A.7 show how the main results vary when using kernel matching instead of nearest neighbour matching estimator. Even if some slight differences are found, the main conclusions remain unchanged. Regarding SMT, the kernel estimator exhibits slightly more positive results on earnings and more negative on occupational mobility and for LMT all results are slightly more positive. However, most of these differences are not statistically significant. Furthermore, when looking at the placebo estimates one year before the starting observation point, the kernel estimates show unbalances. Kernel estimates performs thus less well than the preferred specification in terms of potential bias.

3.3.4. The analysis of the introduction of SMT uses matching and DID analysis

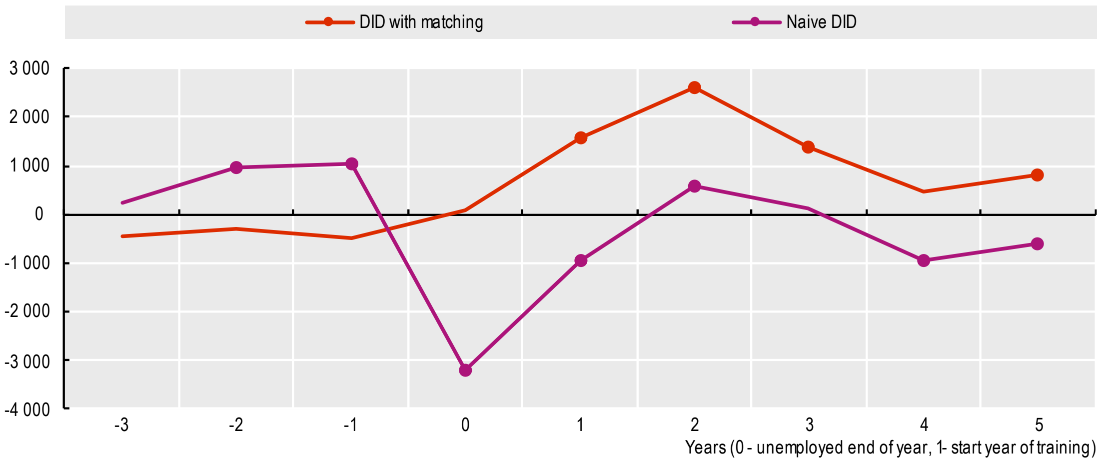
The analysis in the main report of how the introduction of SMT financing impacts on aggregate jobseeker outcomes from training, relative to when only LMT was available uses a combination of matching and DID analysis to analyse impacts. This combination enables both time stable differences between groups to be accounted for and matching allows the incorporation of observed characteristics that may have differential impacts over time.

The naïve DID estimator fails the parallel trends assumption

Figure 3.5 shows that a DID approach without matching the cohorts produces pre-programme outcome trends that are significantly different between the cohorts. This means that the parallel trends assumption is violated. The DID estimator combined with matching rectifies this issue and the parallel trends assumption is now satisfied- the red line is centred around 0 in the pre-programme years (-3 through to 0) and is not statistically significant. This phenomenon is also present when looking at annual months of employment and unemployed, though the data are not presented here for brevity.

Figure 3.5. The matched DID estimator satisfies the parallel trends assumption

Impact on annual earnings, LMT+SS+SMT cohort compared to LMT+SS only



Note: Annual earnings in 2022 earnings terms. All changes are relative to year -4. Series provide additional yearly impact of LMT+SS+SMT cohort compared to LMT+SS only. Matching estimator uses nearest-neighbour propensity score matching results which matches individuals based on several characteristics: duration in unemployment until the point of observation, history in unemployment over the past three years (spells and days), age, gender, marital status, education level and field, Finnish national, municipality type, previous employment, income. Differences statistically significant at 5% confidence level denoted using circular markers.

Source: OECD analysis of Statistics Finland TEM and FOLK datasets

Falsification tests bring into question the validity of the parallel trends assumption

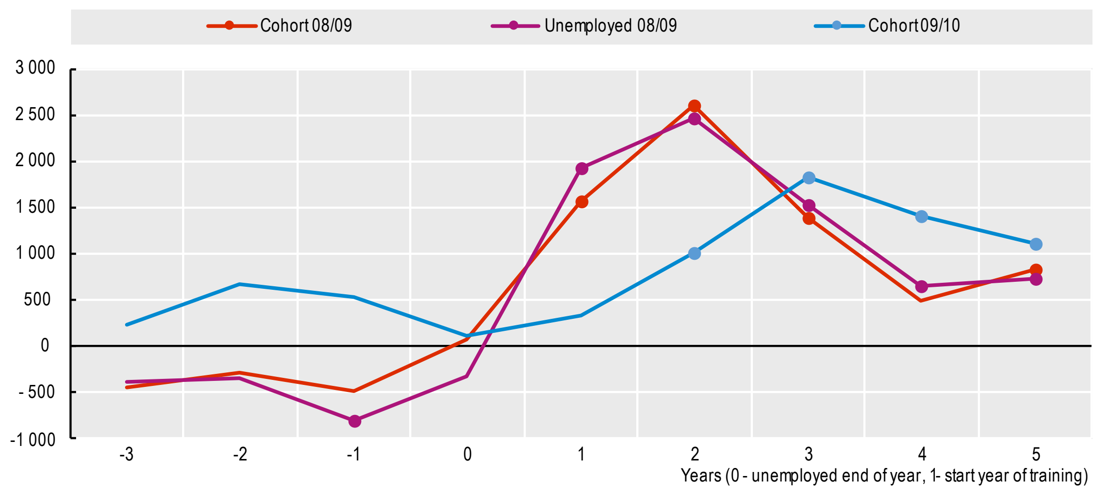
However, it is not possible to conclude that the analysis in Figure 3.5 implies that the LMT+SS+SMT cohort experience greater growth in earnings relative to the LMT+SS only cohort, unless the parallel trends assumption holds over time. Whilst it has been demonstrated that this assumption holds for the pre-programme earnings, it is possible to use falsification tests to determine whether this assumption is plausible in the post-programme period as well.

The first falsification test uses a cohort of individuals that are unemployed in the same years as the LMT+SS/LMT+SS+SMT cohorts, but who receive no training. The expectation should be that there are no significant differences in annual earned income between these two groups. However, Figure 3.6 demonstrates that these cohorts of individuals follow very similar patterns of earnings, suggesting that differences are driven mainly by annual effects of the economic cycle which do change over time, violating the assumption of parallel trends.

Further to this a second falsification test was undertaken using two later training cohorts. It compares the earnings of a cohort of LMT, SS or SMT training participants who were unemployed in 2010, to an earlier cohort of LMT, SS or SMT participants who were unemployed in 2009. Again, Figure 3.6 shows that despite pre-programme years with insignificant differences, later on in the reference period, from year two onwards, significant differences do emerge between the two groups, suggesting that here also effects of the economic cycle play a part (this could equally be due to changing effects of training effectiveness in different years, though previous analysis of LMT has shown that effects are largely similar across different cohort years (Alasalmi et al., 2022[19]).

Figure 3.6. The unemployed without training follow a very similar annual trend

Annual earnings, falsification tests with two unemployed no training cohorts and two later cohorts with training



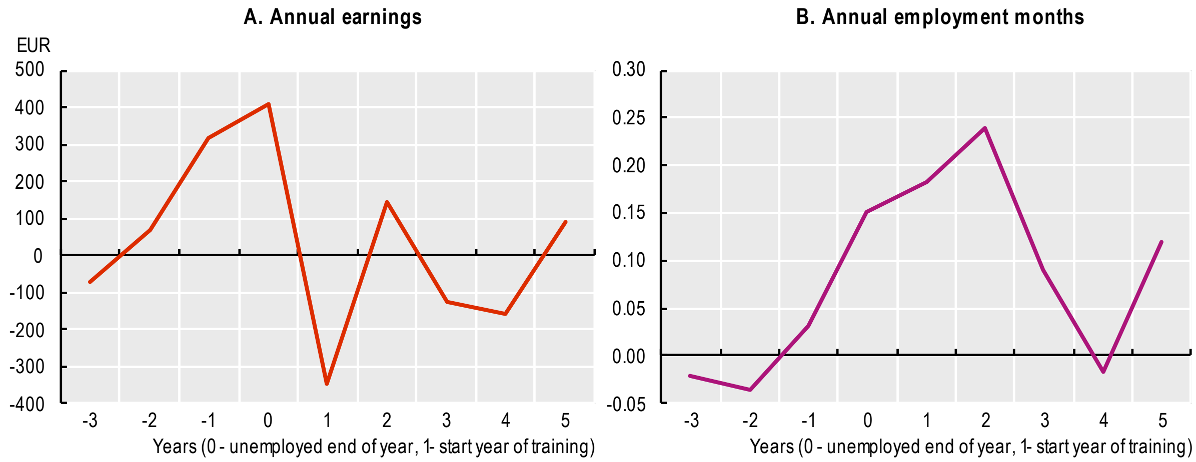
Note: All impacts on annual earnings. Cohort 08/09 refers to the annual earnings of the 2009 LMT+SS+SMT training cohort relative to the 2008 LMT+SS only training cohort. Unemployment 08/09 refers to the annual earning of a group of unemployed individuals in 2009 with no training relative to a similar group of unemployed individuals in 2008 with no training. Cohort 09/10 refers to the annual earnings of the 2010 LMT+SS+SMT training cohort relative to the 2009 LMT+SS+SMT training cohort. Differences statistically significant at 5% confidence level denoted using circular markers.
Source: OECD analysis of Statistics Finland TEM and FOLK datasets.

It is possible to conduct a difference-in-difference-in-difference estimator, which uses a control group to try to remove possible compounding time effects. However, applying this approach to the analysis conducted above is inconclusive. Figure 3.7 demonstrates the estimated impact on earnings and employment for the

LMT+SS+SMT training cohort of unemployed jobseekers in 2009, relative to the LMT+SS only training cohort of jobseekers in 2008, after having subtracted off outcomes for a control group of unemployed individuals in those years who did not participate in any training. The idea is that the control group of unemployed individuals serves as a measure for the year specific economic effects that it is not possible to account for in a simple DID strategy. It is equivalent for earnings of subtracting the blue line from the red line in Figure 3.6. The pattern for annual earnings, in panel A of Figure 3.7 displays a pattern that is a priori expected, in that the impact of earnings is negative in year one. This is expected due to the relatively longer duration of SMT training, compared to LMT/SS training. So the expectation that annual earning are lower can be driven by the greater lock-in period of the SMT training. However, this pattern is not observed in the series for annual employment months in panel B, where months of employment have already started to rise in year one. It is difficult to reconcile this with either the pattern to annual earnings, or to the expectation of the longer lock-in period (which has already been demonstrated in chapter 5 and 6 of the main report (2023[1])). This leads to a conclusion that it is prudent not to interpret these effects as causal. In this analysis, the outcomes for jobseekers without training must serve as a proxy for outcomes of jobseekers with training and this assumption is not directly testable. The interpretation of the charts below means that some caution is therefore necessary.

Figure 3.7. Netting off annual economic effects is inconclusive

Additional outcome impacts of LMT+SS+SMT after subtracting annual effects of economic cycle



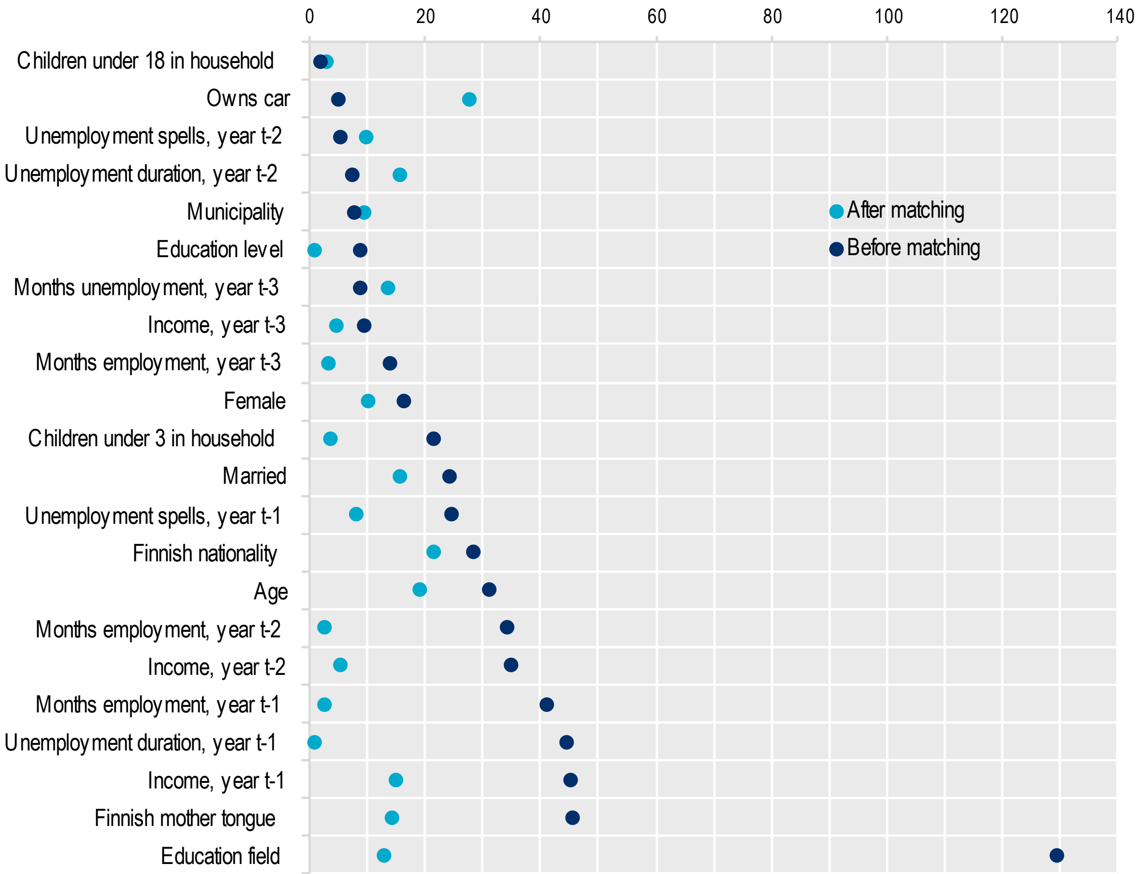
Note: The analysis presents the additional annual earnings and employment months of the LMT+SS+SMT training cohort who were unemployed in 2009, relative to the LMT+SS only training cohort who were unemployed in 2008, after having subtracted off the difference in annual earnings and employment months found for a control group of unemployed individuals without any training in the same years.
Source: OECD analysis of Statistics Finland TEM and FOLK datasets.

Comparing SMT to the Study Subsidy

Similar to the technique used in the individual analysis of LMT and SMT, the comparison of SMT to the study subsidy (SS) in the main report (2023[1]) also uses matching to compare individuals. The cohort sample period is also 2012-14. This matching reduces differences in the underlying characteristics (Figure 3.8) between the groups and lowers overall standardised mean bias by 63% but does not remove all of the differences between the two groups. The pre-programme outcomes of the two groups are insignificantly different from zero (except annual earnings in year zero), suggesting that the most important influences on future outcomes are accounted for, however the degree of remaining imbalance in the characteristics of participants and non-participants means that some caution remains in the interpretation of the overall analysis.

Figure 3.8. Matching reduces the differences between SMT and SS participants

Standardised mean bias SS participants compared to SMT participants

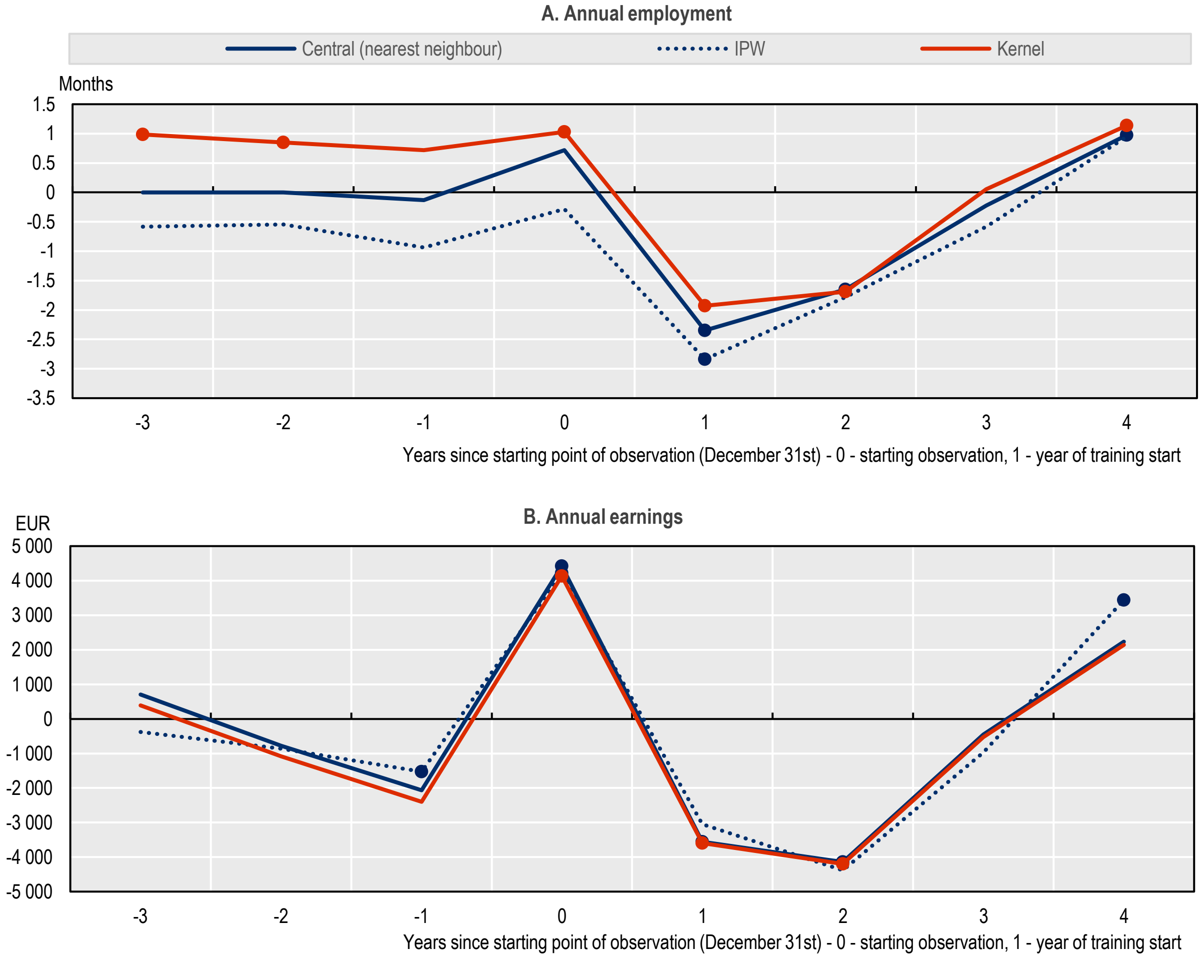


Note: Years are relative to the year the individual is observed unemployed at the end of the year. Standardised differences are calculated as the difference in means between the treatment and control groups for the matching variable divided by the square root of the sums of the variances for that variable.
Source: OECD analysis of Statistics Finland datasets including both FOLK and TEM datasets.

As per the analysis for LMT and SMT, alternative estimators are used to determine how sensitive the results are to estimate used. Kernel matching and inverse-probability weighting are used as the alternative estimators to the nearest-neighbour matching used for the central analysis (2023[1]). Figure 3.9 shows that estimates are of a similar magnitude using the two alternative methods to construct similar participant and non-participant groups. For annual employment, the nearest neighbour estimate has point estimates which balance the pre-programme outcomes most centrally around zero. For annual earnings, there is a closer grouping between the three estimators used. Neither of the two alternative estimators dramatically alters the interpretation derived from the nearest neighbour matching.

Figure 3.9. Estimates are similar across different matching estimators

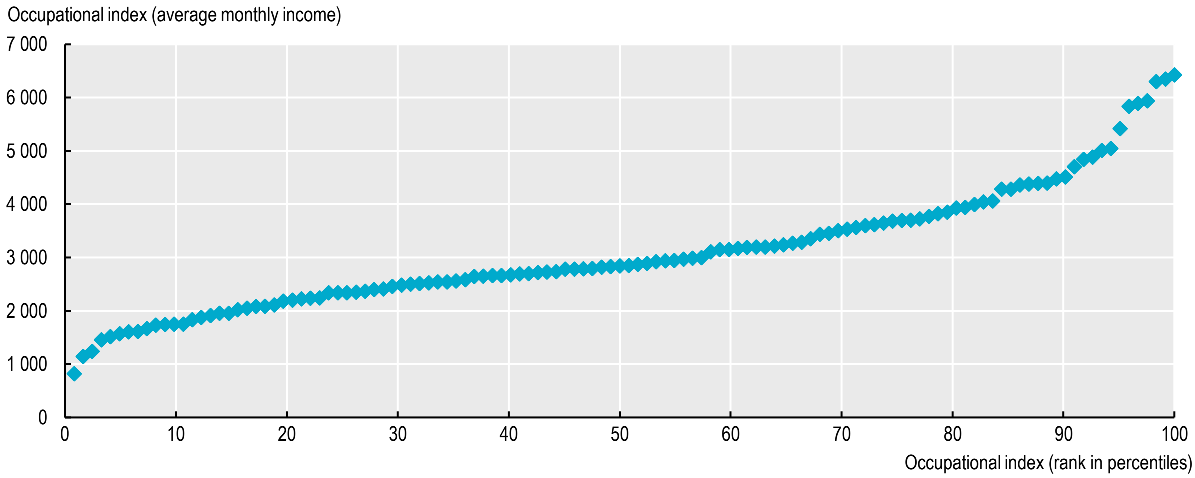
Impact on earnings and employment of SMT training cohort relative to a SS training cohort



Note: Data points shown with a circular marker at significant at the 5% confidence level. IPW = inverse probability weighting estimator. Kernel estimator using Gaussian kernel. Participation probability calculated using a number of characteristics: duration in unemployment until the point of observation, history in unemployment over the past three years (spells and days), age, gender, marital status, education level and field, Finnish national, municipality type, previous employment, income. Estimated participation probabilities are then used to calculate the inverse probability weights and are also used for the nearest neighbour and Kernel matching estimators.
Source: OECD analysis of Statistics Finland datasets including both FOLK and TEM datasets.

Annex 3.A. Additional figures

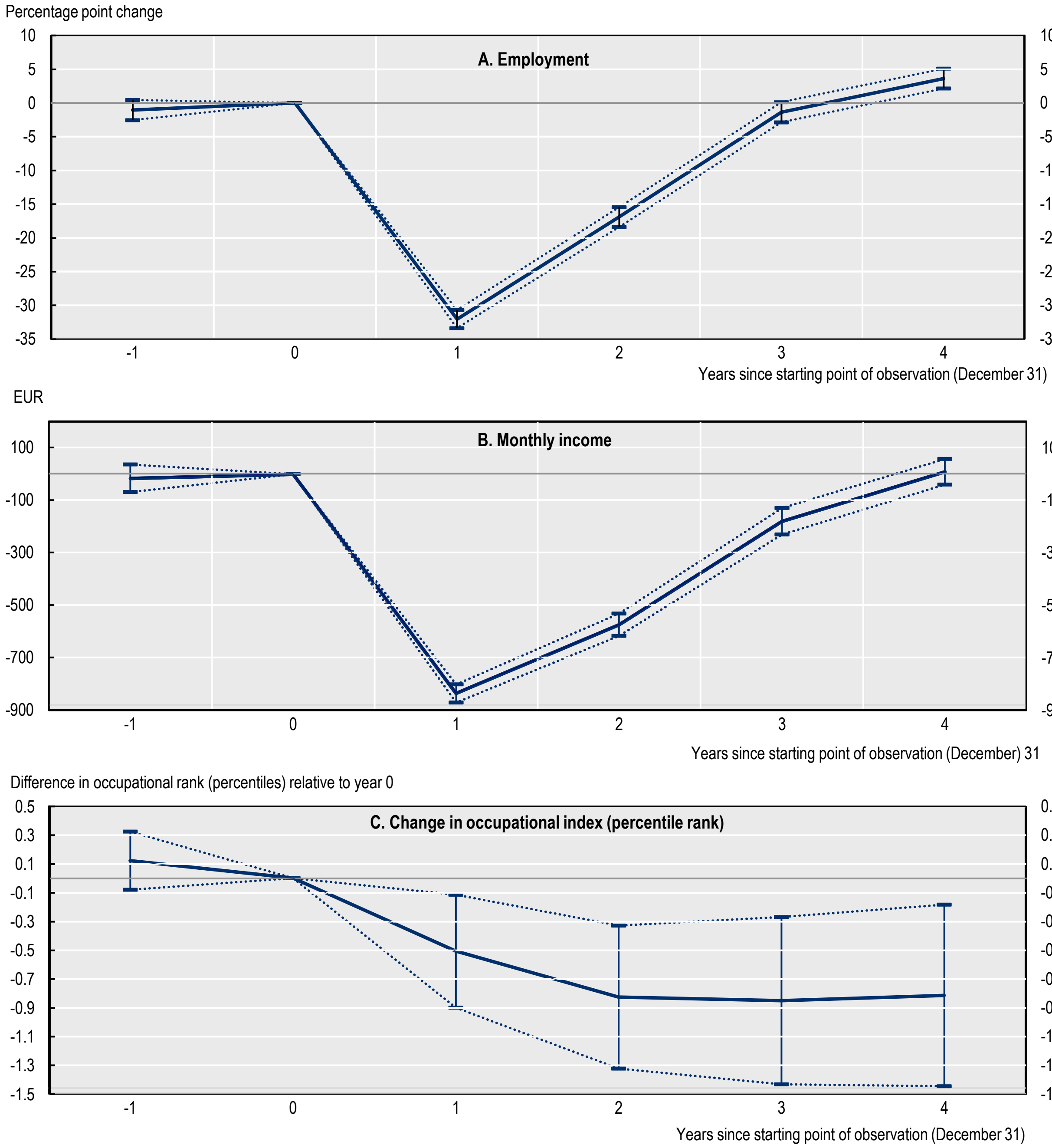
Annex Figure 3.A.1. Relationship between the occupational index in average monthly earnings and the occupational index in percentile rank



Note: The occupational Index expressed as the occupational average monthly earnings is in the vertical axis and the index in percentiles on the horizontal axis. The averages are computed over the 2012-18 period. Each point in the chart represents one of the 122 3-digit occupations that are part of the occupational index in Finland.
Source: OECD calculations based on FOLK datasets. Impact Assessment and Cost-benefit framework.

Annex Figure 3.A.2. FOLK data: SMT has a positive effect on employment and a slightly negative effect on occupational mobility in the long-term

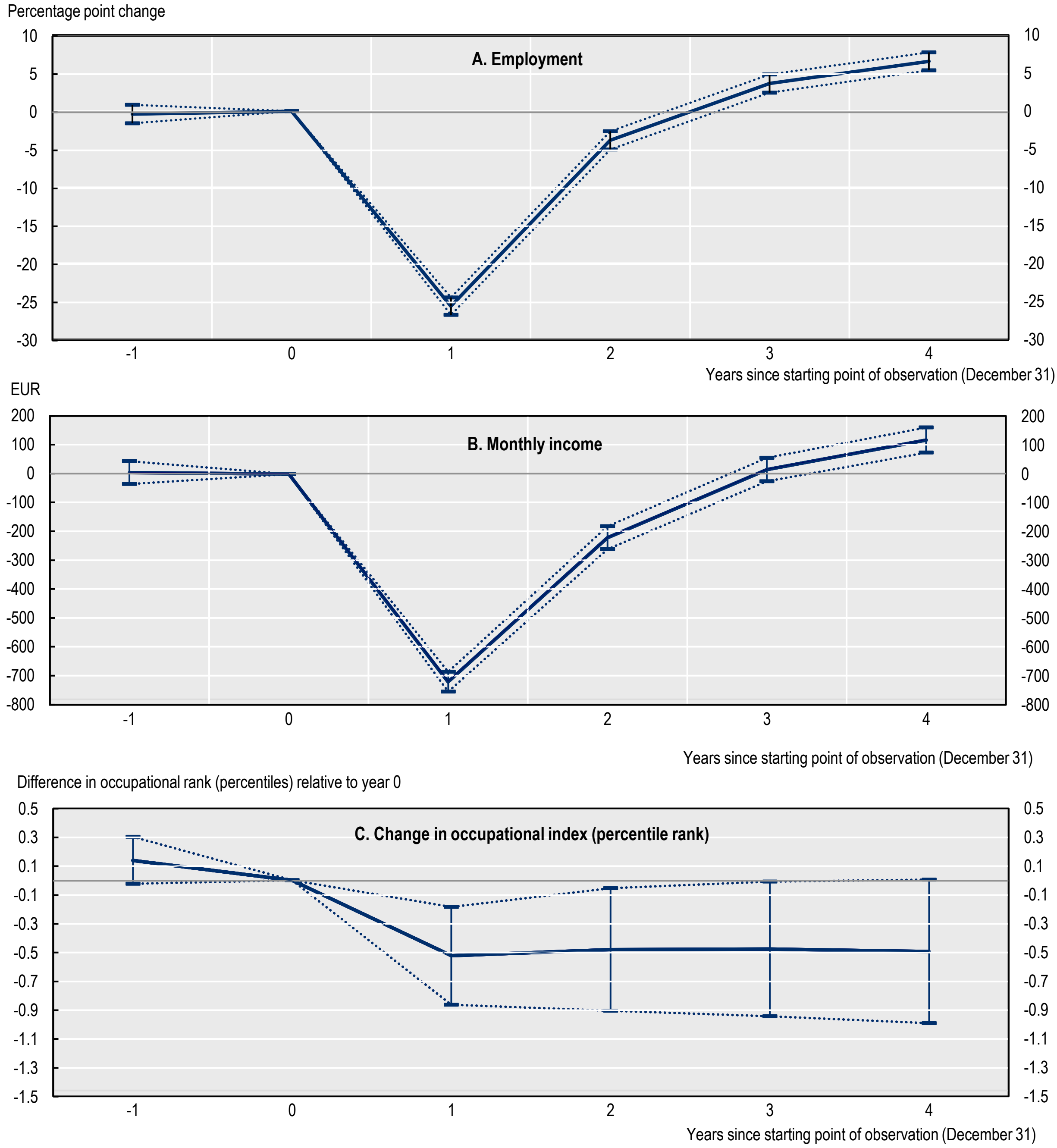
Percentage point change in employment probability (Panel A), change in monthly earnings (Panel B) and change in the occupational index in percentile rank relative to year 0 (Panel C)



Note:: The analysis presents nearest-neighbour propensity score matching results which matches individuals based on a number characteristics: duration in unemployment until the point of observation, history in unemployment over the past two years (spells and days), age, gender, marital status, education level and field, Finnish national, the type of municipality, the type of building, the quarter of registration into unemployment (time trend), as well as previous year employment status, earnings and occupational quality (rank). The confidence intervals are shown at the 5% level of significance and represented by the whiskers delimiting the dotted lines on the charts
Source: OECD calculations based on FOLK datasets.

Annex Figure 3.A.3. FOLK data: LMT has positive effects on both employment and earnings in the long term and no effect on occupational mobility.

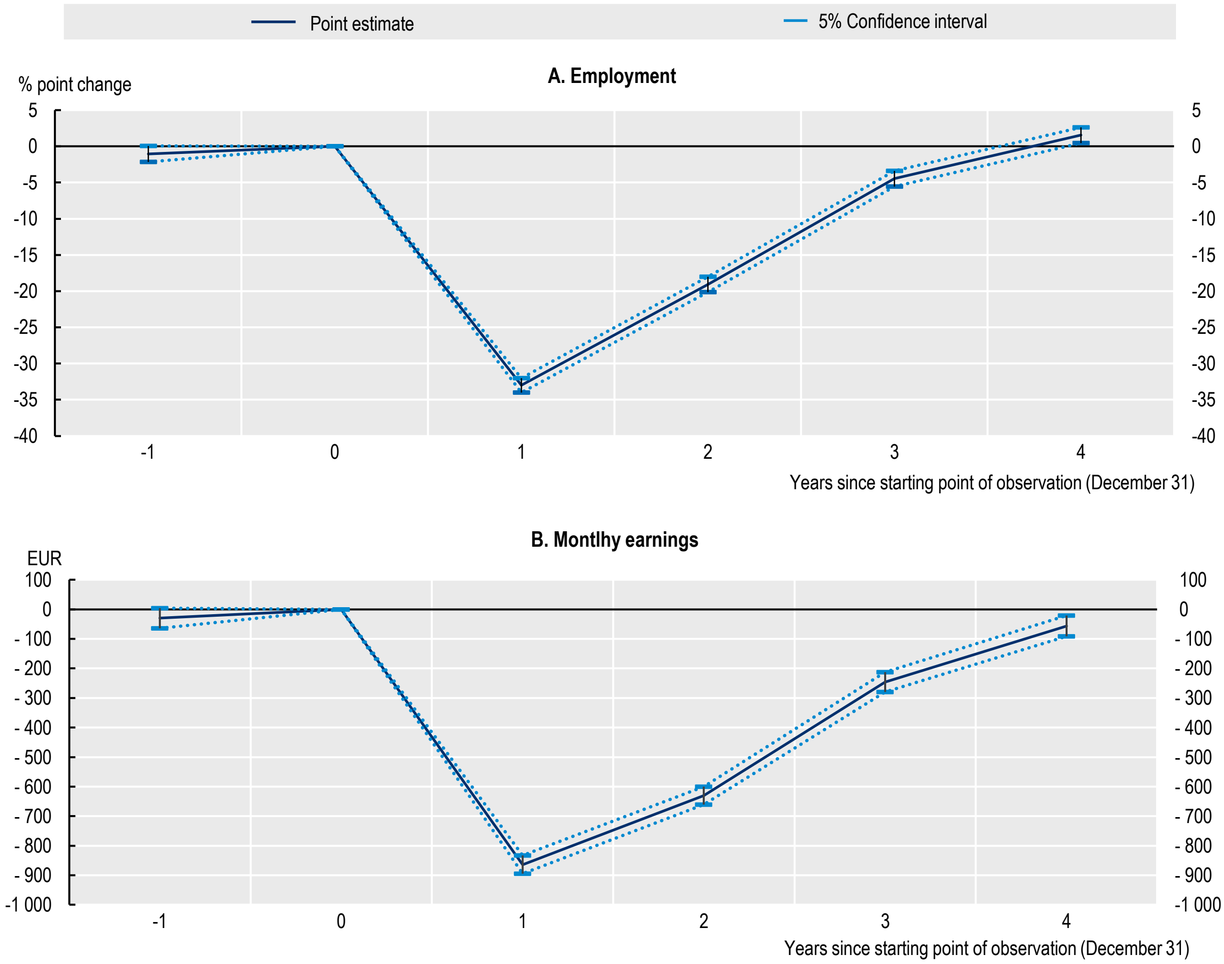
Percentage point change in employment probability (Panel A), change in monthly earnings (Panel B) and change in the occupational index in percentile rank relative to year 0 (Panel C)



Note:: The analysis presents nearest-neighbour propensity score matching results which matches individuals based on a number characteristics: duration in unemployment until the point of observation, history in unemployment over the past two years (spells and days), age, gender, marital status, education level and field, Finnish national, the type of municipality, the type of building, the quarter of registration into unemployment (time trend), as well as previous year employment status, earnings and occupational quality (rank). The confidence intervals are shown at the 5% level of significance and represented by the whiskers delimiting the dotted lines on the charts
Source: OECD calculations based on FOLK datasets.

Annex Figure 3.A.4. Change in the sample: The effect of SMT in the long run is slightly positive for employment and slightly negative for earnings

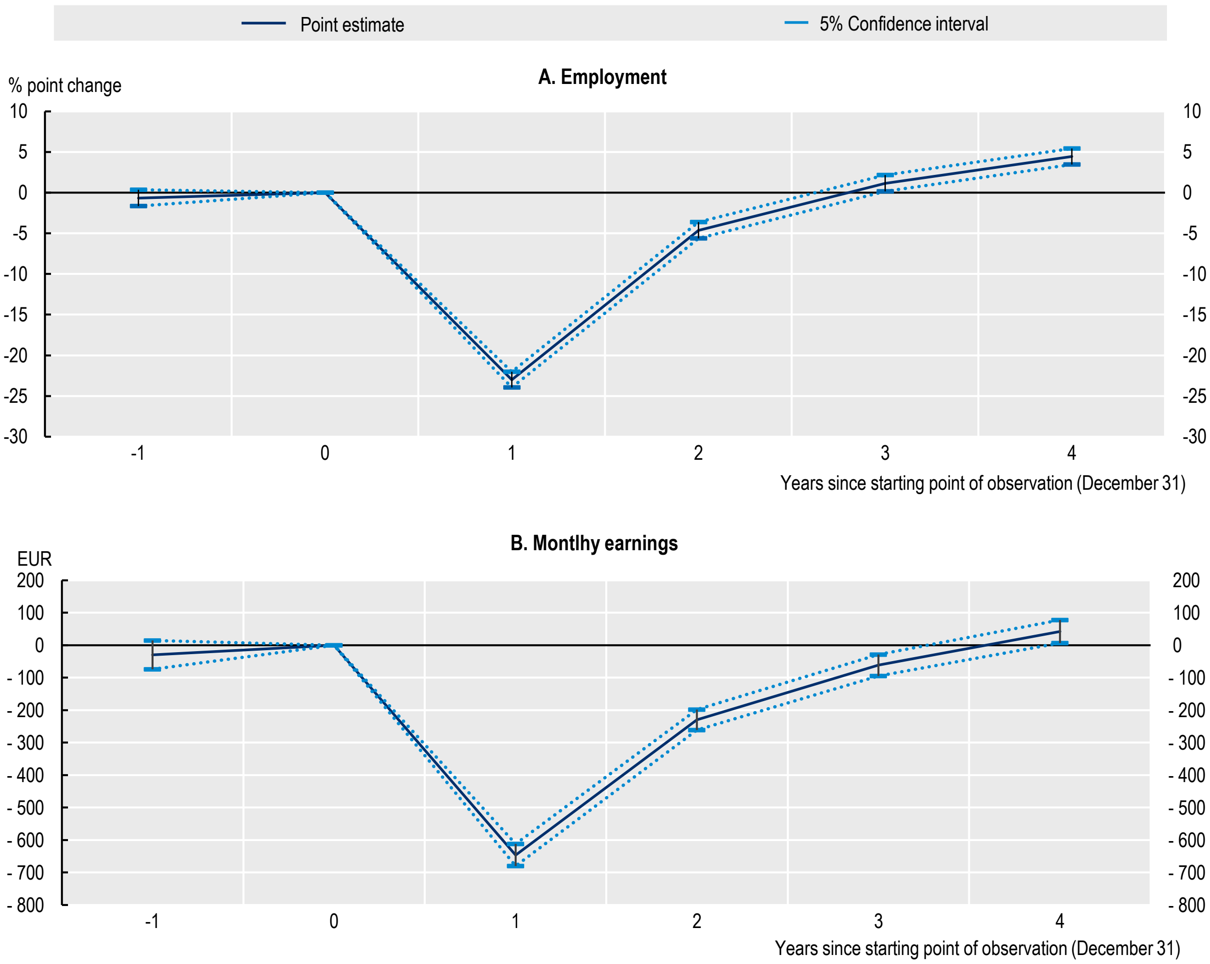
Percentage point change in employment probability (Panel A) and change in monthly earnings (Panel B)



Note: The analysis presents nearest-neighbour propensity score matching results which matches individuals based on a number characteristics: duration in unemployment until the point of observation, history in unemployment over the past two years (spells and days), age, gender, marital status, education level and field, Finnish national, the type of municipality, the type of building, the quarter of registration into unemployment (time trend), as well as previous year employment status, earnings and occupational quality (rank). The confidence intervals are shown at the 5% level of significance and represented by the whiskers delimiting the dotted lines on the charts
Source: OECD calculations based on FOLK and TEM datasets.

Annex Figure 3.A.5. Change in the sample: The effects of LMT in the long run are positive for both employment and earnings

Percentage point change in employment probability (Panel A) and change in monthly earnings (Panel B)

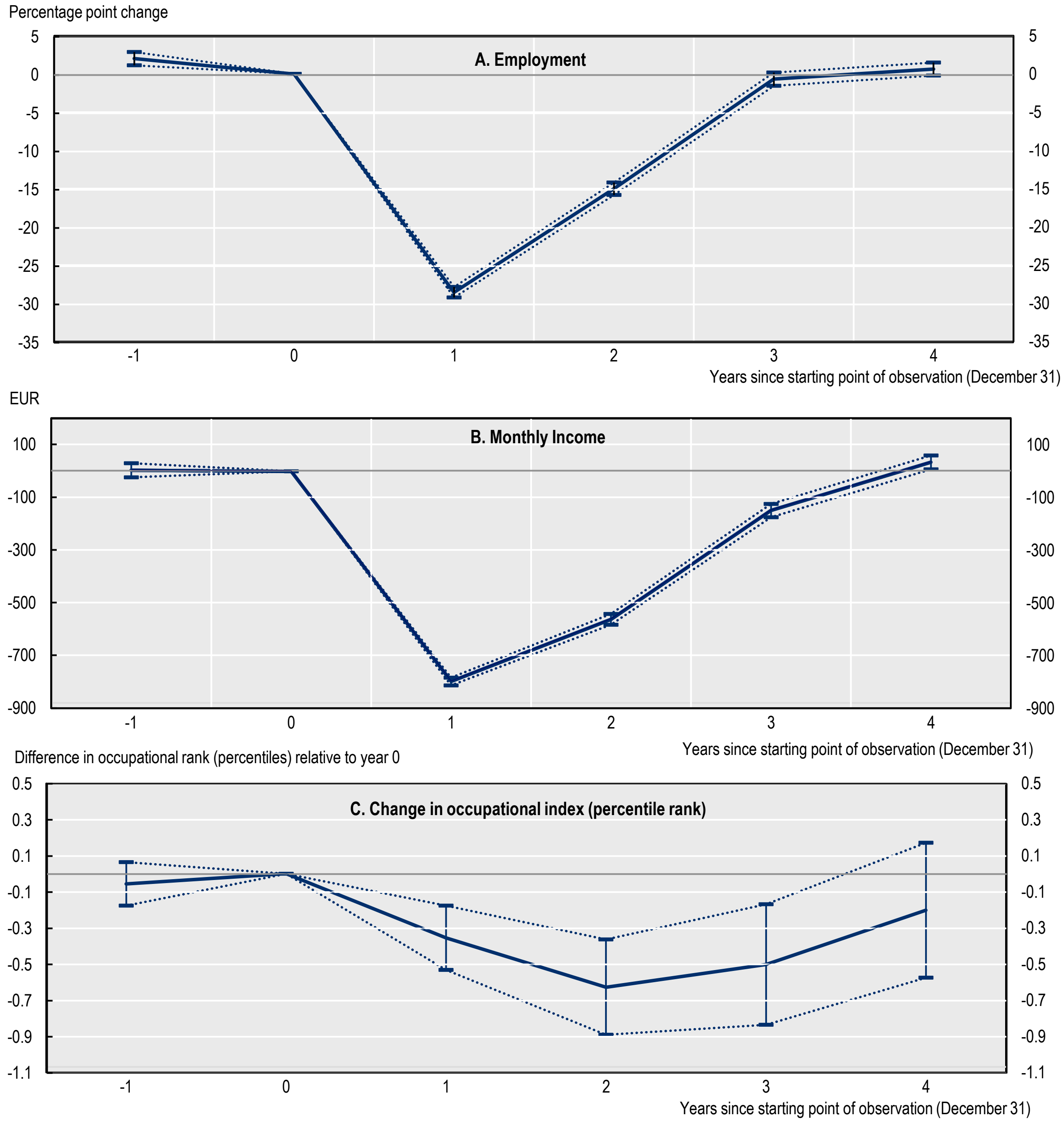


Note: The analysis presents nearest-neighbour propensity score matching results which matches individuals based on a number characteristics: duration in unemployment until the point of observation, history in unemployment over the past two years (spells and days), age, gender, marital status, education level and field, Finnish national, the type of municipality, the type of building, the quarter of registration into unemployment (time trend), as well as previous year employment status, earnings and occupational quality (rank). The confidence intervals are shown at the 5% level of significance and represented by the whiskers delimiting the dotted lines on the charts

Source: OECD calculations based on FOLK and TEM datasets.

Annex Figure 3.A.6. Kernel PSM: The effect of SMT on employment and earnings is slightly positive and there is no effect on occupational mobility

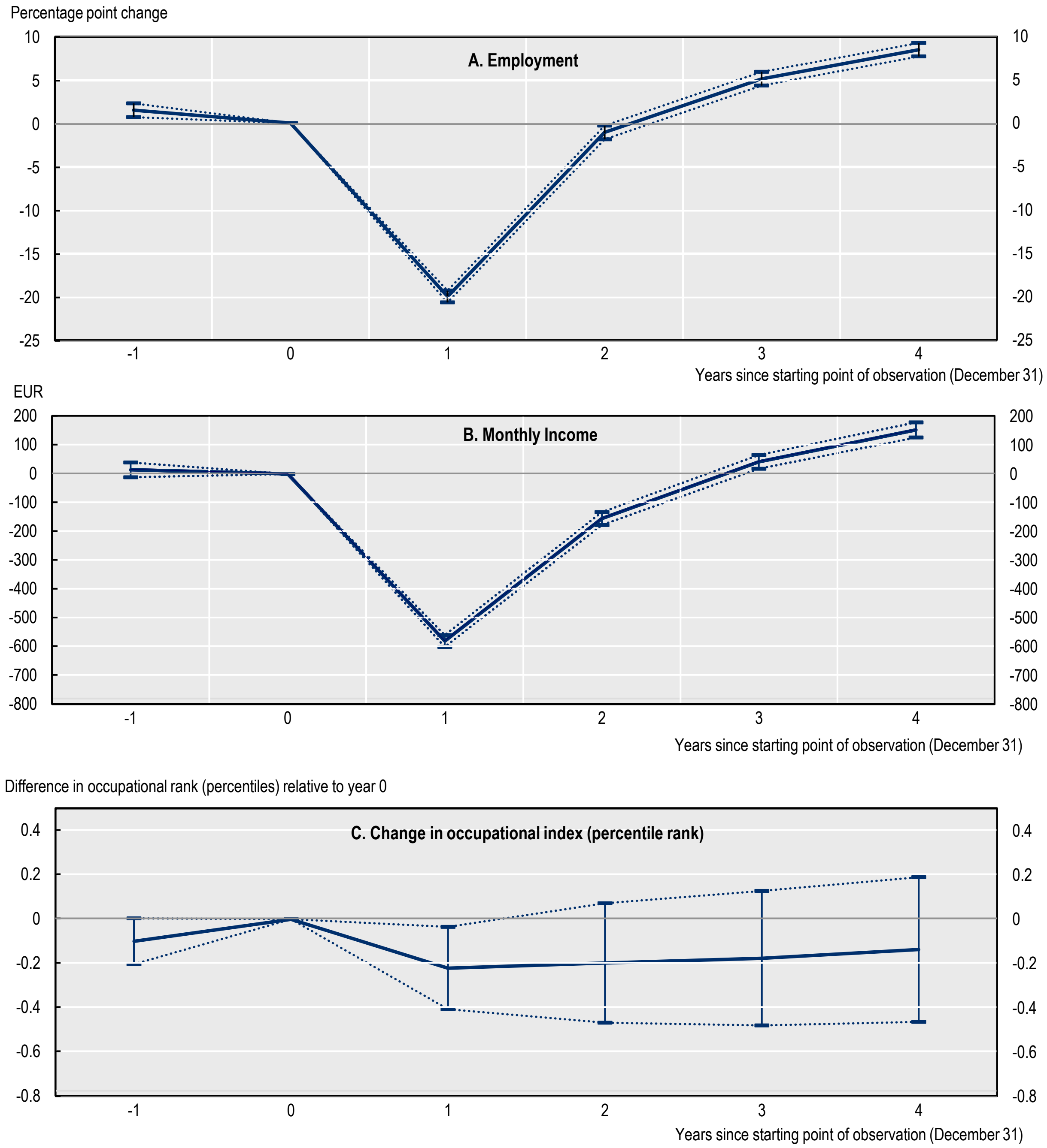
Percentage point change in employment probability (Panel A), change in monthly earnings (Panel B) and change in the occupational index in percentile rank relative to year 0 (Panel C)



Note:: The analysis presents kernel propensity score matching results which matches individuals based on a number characteristics: duration in unemployment until the point of observation, history in unemployment over the past two years (spells and days), age, gender, marital status, education level and field, Finnish national, the type of municipality, the type of building, the quarter of registration into unemployment (time trend), as well as previous year employment status, earnings and occupational quality (rank). The confidence intervals are shown at the 5% level of significance and represented by the whiskers delimiting the dotted lines on the charts
Source: OECD calculations based on FOLK and TEM datasets.

Annex Figure 3.A.7. Kernel PSM: LMT has positive effects on both employment and earnings in the long term and no effect on occupational mobility

Percentage point change in employment probability (Panel A), change in monthly earnings (Panel B) and change the occupational index in percentile rank relative to year 0 (Panel C)



Note: The analysis presents kernel propensity score matching results which matches individuals based on a number characteristics: duration in unemployment until the point of observation, history in unemployment over the past two years (spells and days), age, gender, marital status, education level and field, Finnish national, the type of municipality, the type of building, the quarter of registration into unemployment (time trend), as well as previous year employment status, earnings and occupational quality (rank). The confidence intervals are shown at the 5% level of significance and represented by the whiskers delimiting the dotted lines on the charts
Source: OECD calculations based on FOLK and TEM datasets.

Annex Table 3.A.1. SMT Bachelors and Masters students have higher previous income and are less likely to live in a rural area

SMT participant characteristics by type of course studied

	Vocational Upper Secondary	Further Education	Polytechnic Bachelors	University Bachelors	Masters
Age	35.99	38.52	34.29	36.93	37.22
Children	1.00	0.96	0.89	0.89	0.85
Months of Unemployment	3.29	3.82	3.32	3.30	3.23
Months of Employment	2.51	3.17	3.85	3.99	4.42
Earned Income	3 700	5 300	7 000	8 900	10 700
Lower Work	17%	20%	18%	11%	11%
Owns Car	42%	49%	45%	37%	43%
Finnish Language	69%	80%	91%	93%	88%
Finnish Citizen	86%	93%	97%	99%	96%
Upper Secondary Education	48%	53%	61%	48%	3%
Tertiary Education	10%	13%	27%	44%	94%
Single	59%	58%	63%	61%	56%
Rural	16%	18%	11%	9%	8%
Female	61%	64%	61%	63%	66%
Single Mother	10%	11%	12%	14%	15%
Single Father	16%	14%	17%	17%	12%
Married Mother	38%	37%	35%	34%	34%

Note: Includes SMT records from all years 2010 to 2018, matched to both education and FOLK basic data. Looks at 1st instance of SMT recorded on the data per individual. Around 13% of SMT records do not match to education and/or individual characteristics datasets. All characteristics come from the FOLK Basic data and relate to annual records in the year before the start year of the SMT spell identified. For example, Months of Unemployment/Employment, relate to the number of months in the year before SMT entry.
Source: OECD analysis of TEM, FOLK and OPISK datasets, accessed via Statistics Finland

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. [15]

Alasalmi, J. et al. (2022), *Ammatillisen työvoimakoulutuksen toimivuuden ja vaikutusten arviointi (Evaluation of the effectiveness and impact of vocational education training)*, Prime Minister’s Office Helsinki 2022. [19]

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. [18]

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. [13]

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. [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. [10]

Cheung, M. et al. (2019), “Does Job Search Assistance Reduce Unemployment? Experimental Evidence on Displacement Effects and Mechanisms”, *SSRN Electronic Journal*, https://doi.org/10.2139/SSRN.3515935. [2]

Desiere, S., K. Langenbucher and L. Struyven (2019), *Statistical profiling in public employment services: An international comparison*, OECD Social, Employment and Migration Working Papers 224, OECD Publishing. [3]

DWP (2018), “Weekly Work Search Review Trial”, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/721910/weekly-work-search-review-trial.pdf (accessed on 26 November 2021). [5]

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. [16]

Gertler, P. et al. (2016), *Impact Evaluation in Practice, Second Edition*, Washington, DC: Inter-American Development Bank and World Bank, https://doi.org/10.1596/978-1-4648-0779-4. [6]

Heckman, J. et al. (1998), “Characterizing Selection Bias Using Experimental Data”, *Econometrica*, Vol. 66/5, p. 1017, https://doi.org/10.2307/2999630. [12]

Heckman, J. et al. (1998), “Characterizing Selection Bias Using Experimental Data”, *Econometrica*, Vol. 66/5, p. 1017, https://doi.org/10.2307/2999630. [14]

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. [17]

Laporšek, S. et al. (2021), “Winners and losers after 25 years of transition: Decreasing wage inequality in Slovenia”, *Economic Systems*, Vol. 45/2, p. 100856, https://doi.org/10.1016/j.ecosys.2021.100856. [7]

Neyman, J. (1923), “On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.”, *Statistical Science*, Vol. 5/4. [8]

OECD (2023), *Evaluation of active labour market policies in Finland*, OECD Publishing, Paris. [1]

OECD (2022), *Impact evaluation of the digital tool for employment counsellors in Spain: SEND@*, OECD Publishing, Paris. [4]

Rubin, D. (2006), *Matched Sampling for Causal Effects*, Cambridge University Press, . Cambridge, England. [9]

Notes

[1] Taika research catalogue available at: https://taika.stat.fi/