First published: 1 November 2013
Last updated: 2 September 2013

Active Labour Market Programme Participation for Unemployment Insurance Recipients: protocol for a systematic review

Trine Filges, Lars Pico Geerdsen, Geir Smedslund, Anne-Sofie Due Knudsen, Anne-Marie Klint Jørgensen

PROTOCOL

THE CAMPBELL COLLABORATION

Contact details of lead author:
Trine Filges
SFI / The Danish National Centre for Social Research
Herluf Trolles Gade 11
Copenhagen K
DK- 1052 **DENMARK**
Email: tif@sfi.dk

Anne-Sofie Due Knudsen and Anne-Marie Klint Jørgensen are also employed by SFI.

Lars Pico Geerdsen is based at the Kraks Fond Byforskning, Copenhagen, Denmark

Geir Smedslund is based at the Norwegian Knowledge Centre for the Health Services, Oslo, Norway

Table of contents

1 Background

1.1 DESCRIPTION OF THE CONDITION

During the 1990s, many countries introduced Active Labour Market Programmes (ALMPs) in an effort to reduce unemployment. In countries such as Australia, USA, Denmark, Sweden, England and Switzerland, participation in an active labour market programme is required if an unemployed individual is to continue receiving benefits (Gerfin & Lecher, 2002; Geerdsen, 2003). Typically, compulsory programme participation is required after the individual has received unemployment benefits for a certain period of time.

The purpose of making benefit payments conditional on participation in ALMPs is twofold. Firstly, participation in ALMPs may improve the participants' qualifications and so allow their reintroduction into the labour market. Secondly, the compulsory aspect may provide an incentive for unemployed individuals to look for and return to work *prior* to programme participation (Black, Smith, Berger & Brett, 2003; Jackman, 1994; Hansen & Tranæs, 1999). This is sometimes referred to as the 'threat effect', and a systematic review of this effect occurring prior to participation in compulsory labour market programmes is currently in progress (Geerdsen, Bjørn, Filges & Jensen, 2011).

We will focus on research on the outcome of programme participation (i.e. effects *during* and *after* programme participation,Heckman, Lalonde & Smith, 1999; Martin & Grubb, 2001). The effects of ALMP participation on job-finding rates are typically composed of two separate effects: a lock-in effect and a post-programme effect. The lock-in effect refers to the period of participation in a programme. During this period, job-search intensity may be lowered because there is less time to search for a job, and participants may want to complete an on-going skill-enhancing activity; hence the lock-in effect. The post-programme effect refers to the period after participation in a programme. If the ALMP has increased the individual's employability, a rise in the job-finding rate is expected. The combination of these

two effects consequently determines the net effects of ALMP participation on unemployment duration.

1.2 DESCRIPTION OF THE INTERVENTION

The intervention is ALMP participation by those in receipt of unemployment insurance benefits. A variety of ALMPs exist across different countries. They can consist of job search assistance, training, education, subsidized work, and similar programmes. Some of the programmes (e.g. subsidized work, training and education) demand full-time participation over a long time period (e.g. several months), while other programmes (e.g. job search assistance and education) are part-time and have a short duration (e.g. few days/weeks). It is possible to classify these programmes into a set of four core categories. The categories we will use correspond to classifications that have been suggested and used by the OECD and Eurostat (OECD, 2004 and Eurostat, 2005). These are described below in detail, under 'Types of interventions'. Studies that have included individuals receiving other types of unemployment benefits as well as individuals receiving unemployment insurance benefits will be included in this review if more than 60 per cent of those included were receiving unemployment insurance benefits.

1.3 HOW THE INTERVENTION MIGHT WORK

Active labour market programmes were adopted by most advanced countries during the 1990s (Gerfin & Lechner, 2002). The declared purpose of such policies is to protect workers who are exposed to negative employment shocks due to changing market conditions. Programmes that involve subsidized work, training and education are designed to reduce skill loss during extended periods of unemployment and to redirect the skills of those who are left without work as a result of new technology or increased international trade (Kluve et al., 2007). The introduction of ALMPs is thus often motivated by the need to upgrade the skills of those suffering long-term unemployment to improve their productivity and, subsequently, their employability. If participation in an ALMP increases the individual's employability, a rise in the job-finding rate is to be expected; however, the increased human capital may result in higher reservation wages[1], effectively

[1] The minimum wage at which a job offer is acceptable.

offsetting the positive employment effect (Filges, Kennes, Larsen & Tranæs, 2011; Mortensen, 1987). Moreover, some programmes may stigmatize workers in the view of potential employers. Programmes associated with participants having poor employment prospects (e.g. the long term unemployed) may carry a stigma. Because of asymmetric information (a situation where there is imperfect knowledge in which one party has different information to another), employers cannot know the productivity of new workers, some of whom they might hire from the pool of the unemployed. Prospective employers might then perceive participants in such employment programmes as low productivity workers or as workers with a tenuous labour market attachment (Kluve, Lehmann & Schmidt, 1999; Kluve et al., 2007).

Finally, some ALMPs are designed to encourage the unemployed to return to work and may increase the efficiency of the matching process. For example, job search assistance is expected to increase the search intensity of participants and therefore directly enhance the matching efficiency between vacancies and the unemployed (Pissarides, 2000).

1.4 WHY IT IS IMPORTANT TO DO THIS REVIEW

There is currently considerable political interest in reducing levels of unemployment, and the use of ALMPs as a means of achieving this goal has been advocated (Filges et al., 2011; Kluve et al., 2007). At the same time, ALMPs has been heavily criticized for lack of effectiveness.

Several papers summarise the effect of ALMP (Heckman et al., 1999; Kluve, 2006; Kluve & Schmidt, 2002; Martin, 2000; OECD, 2004; Card, Kluve & Weber, 2010; Martin & Grubb, 2001). However, none are systematic in their search of relevant literature and none provide a synthesis of the magnitude of the effect size, although Kluve (2006) and Card et al. (2010) offer a meta-analysis based on vote counting and an analysis of different covariates' contribution to the probability of obtaining a significant positive, significant negative or non-significant effect.

The effect of active labour market programmes for unemployed people receiving other kinds of unemployment benefits is reviewed in the Campbell Systematic Review "Work programmes for welfare recipients" (Smedslund et al., 2006) where the objective is to estimate the effects of work programmes on welfare recipients' employment and economic self-sufficiency. Individuals who are entitled to

unemployment insurance benefits or who have pensions of any kind are, however, excluded in Smedslund et al. (2006).

There is currently no published systematic review on the effect of ALMP participation in unemployed individuals receiving unemployment insurance benefits..

2 Objective of the review

The objective of this systematic review is to evaluate the effectiveness of ALMP participation on employment status for unemployment insurance recipients.

3 Methods

3.1 CRITERIA FOR CONSIDERING STUDIES FOR THIS REVIEW

3.1.1 Types of studies

The study designs eligible for inclusion are:

- Controlled trials:
  - RCTs - randomized controlled trials
  - QRCTs - quasi-randomized controlled trials where participants are allocated by, for example, alternate allocation, participant's birth date, date, case number or alphabetically
  - NRCTs - non-randomized controlled trials where participants are allocated by other actions controlled by the researcher

- Non-randomized studies (NRS) where allocation is not controlled by the researcher and two or more groups of participants are compared. Participants are allocated by, for example, time differences, location differences, decision makers, policy rules or participant preferences.

We will include study designs that use a well-defined control group. The main control or comparison condition is ordinary (passive) unemployment insurance benefits or the usual services available to unemployment insurance recipients (that are not ALMPs).

Randomised controlled trials that focus on the effectiveness of such social and labour market policies are relatively new in the field. Studies of the effect of ALMP participation are, therefore, typically estimated on observational data, often

obtained from administrative registers or by questionnaires. Studies that use different data sources for their treatment and control groups will not be included in this review.

We will only include studies that use individual micro-data. We will exclude studies that rely on regional or national time series data, even though micro-econometric estimates of individual treatment effects merely provide partial information about the full impact of ALMP (Calmfors, 1994; Calmfors, 1995).

The micro economic literature disregards any deadweight loss and substitution effects[2], as well as any productivity and competition effects. However, reliable empirical evidence which considers all direct and indirect effects on programme participants and on workers not targeted by the intervention is very difficult to generate. At the aggregate level, expenditures for ALMP tend to be high in times of economic recession: this two-way causality between policy measures and outcomes makes it very difficult to assess the impact of the former on the latter and reliable evidence from macro studies is limited. As Heckman et al. (1999) emphasize, accounting for *general equilibrium effects*[3] in a convincing way generally requires the construction of a structural model of the labour market. However, the difficulty of assembling all behavioural parameters for a structural general equilibrium model is substantial, and the conclusions from these models remain controversial, so that their relative value compared to the more traditional 'treatment effect' evaluations continues to be an open research question (Smith, 2000a, 2000b).

3.1.2 **Types of participants**

Eligible participants are unemployed individuals who receive unemployment insurance benefits. The International Labour Office (ILO) definition of an unemployed individual is a person, male or female, aged 15-74, without a job who is available for work and either has searched for work in the past four weeks or is available to start work within two weeks and/or is waiting to start a job already obtained (ILO, 1990); however, different countries may apply different definitions of an unemployed individual, see for example Statistics Denmark (2009). The eligibility rules of unemployment insurance benefits differ between countries. We will exclude individuals receiving other types of benefits such as social assistance

[2] The **deadweight loss** is defined as the hirings from the target group that would have occurred also in the absence of the programme. The **substitution effect** is defined as the extent to which jobs created for a certain category of workers simply replace jobs for other categories, because relative wage costs are changed.

[3] All direct and indirect effects on programme participants and on workers not targeted by the intervention and interactions with other policies.

benefits or benefits not related to being unemployed. In most OECD countries a secondary benefit is available for those who have exhausted regular unemployment insurance benefits. These are known as 'social assistance benefits'. Unlike unemployment insurance benefits, social assistance benefits are generally means-tested, pay a lower level of benefit, and have no time limit. Studies including a mix of individuals receiving unemployment insurance benefits and other individuals receiving social assistance benefits and/or other types of benefits will only be included if more than 60 per cent of the included individuals receive unemployment insurance benefits.

3.1.3 **Types of interventions**

The intervention is participation in ALMP.

ALMPs can include a wide range of activities as listed below. ALMPs typically apply to unemployment insurance beneficiaries and (if different) employable social assistance beneficiaries[4], but similar principles are increasingly being applied to lone-parent and disability beneficiaries[5]. In this review, ALMPs will be understood in the narrow sense of training or employment measures for the unemployed receiving unemployment insurance benefits.

A variety of ALMP programmes exists across countries, which may be classified into four core categories. In this review, we will adopt the classifications suggested and used by the OECD and Eurostat (OECD, 2004; Eurostat, 2005):

A. The first programme type, **(labour market) training**, encompasses measures such as classroom training, on-the-job training and work experience. The training can either provide a more general education (as with language courses, or basic computer courses) or specific vocational skills (as with advanced computer courses or courses providing technical or manufacturing skills). Their main objective is to develop the productivity and employability of the participants and to enhance human capital by increasing skills. Training programmes constitute the "classic" component of ALMP.

[4] In most OECD countries, a secondary benefit (known as social assistance benefit) is available for those who have exhausted regular unemployment insurance benefits (OECD, 2007).

[5] In the US, disability benefit is designed to provide income supplement to people who are physically restricted in their ability to be employed because of a notable, usually physical, disability (CBO, 2010), whereas in Denmark the disability may be both physical and mental (Høgelund & Holm, 2005). Disability benefits can be supplied on either a temporary or permanent basis, usually directly correlated to whether the person's disability is temporary or permanent.

B. **Private sector programmes** are those aimed at creating incentives to alter employer and/or worker behaviour in relation to private sector employment. Wage subsidies are the most commonly used measure in this category. The objective of subsidies is to encourage employers to hire new workers or to maintain jobs that would otherwise be broken up. These can either be direct wage subsidies to employers, or financial incentives that are offered to workers for a limited period of time. The use of self-employment grants form another type of subsidized private sector employment: these grants may be offered to participants who start their own business, sometimes along with advisory support for a fixed period of time (OECD, 2004; Eurostat, 2005).
C. In contrast to subsidies in the private sector, the third programme type, **direct employment programmes in the public sector,** focuses on the direct creation and provision of public works or other activities that produce public goods or services. These measures are mainly targeted at the most disadvantaged individuals, pursuing the aim of keeping them in contact with the labour market and precluding the loss of human capital during a period of unemployment. The created jobs are, nevertheless, often additionally generated and at a distance from the ordinary labour market.
D. The fourth type of programme, **Job search assistance,** encompasses measures aimed at enhancing job search efficiency. The services included are job-search courses and related forms of intensified counselling for those who have difficulty finding employment. The public employment services (PES) often target the disadvantaged and long-term unemployed, whereas private services may focus on the more privileged employees and white-collar workers. These programmes are usually the least expensive.

Programmes that only consist of monitoring (such as carrying out surveillance of the search activities of the unemployed) will not be included. Specialized types of ALMPs targeting only particular groups (such as specialized youth programmes, vocational rehabilitation, sheltered work programmes or wage subsidies for individuals with physical, mental or social disabilities) will be excluded.

3.1.4 **Types of outcomes**

The primary outcome is exits from the unemployment insurance system and into employment[6]. Studies which only examine exits to other destinations, such as other types of social benefits or non-employment, will not be included. Studies that do not distinguish between destinations (i.e. between exits to employment and exits to other destinations such as other types of social benefit or non-employment) will be analysed separately.

In addition to the primary outcome, we will consider secondary outcomes that are relevant to the impact ALMP has on the duration of employment and on income, in order to obtain a clearer picture of the effect the ALMPs have on the quality of the job obtained. If the duration of employment is short or the wage is low, this could indicate that the ALMP forces participants to find jobs that do not match their qualifications, with the consequence that they may return to unemployment quickly.

Primary outcomes refer to employment status:

a) exit rate from unemployment to employment (= work with standard wages and which anyone can apply for)
b) proportion employed (= proportion of participants who have obtained work with standard wages and which anyone can apply for)
c) duration until employment (= work with standard wages and which anyone can apply for)

Secondary outcomes

a) duration of first employment spell post-intervention
b) re-employment wage

3.2 SEARCH METHODS FOR IDENTIFICATION OF STUDIES

3.2.1 **Electronic searches**

Relevant studies will be identified through electronic searches of bibliographic databases, research networks, government policy databanks and internet search engines. No language or date restrictions will be applied to the searches. The following databases will be searched:

[6] When an unemployed person receiving unemployment insurance benefit leaves the unemployment insurance system (e.g. has found a job, withdraws from the labour force, exhausts the benefit period and receives other types of social benefits etc.) there is a tradition in the economics literature for this to be termed an 'exit'.

Business Source Elite

EconLit

PsycINFO

SocIndex

Science Citation Index

Social Science Citation Index

The Cochrane Library (*Cochrane reviews, other reviews*)

International Bibliography of the Social Sciences

IDEAS/Economist Online[7]

3.2.2 **Search terms**

The search strategy for Business Source Elite is listed below. The strategy will be modified for the other databases.

S1 TI (((active labo#r policy) or (active labo#r policies))) or AB (((active labo#r policy) or (active labo#r policies)))

S2 TI ((Labo#r market program*) or LMP or ALMP) or AB ((Labo#r market program*) or LMP or ALMP)

S3 TI ((job or jobs or work or employment*) n1 (help or assistance or program* or training or incentive or scheme* or counsel* or course* or initiative* or experience* or experiment* or training or support*)) or AB ((job or jobs or work or employment*) n1 (help or assistance or program* or training* or incentive or scheme* or counsel* or course* or initiative* or experience* or experiment* or training or support*))

S4 TI (un-employ* or unemploy*) or AB (un-employ* or unemploy*)

S5 3 and s4

S6 TI replac* N1 scheme* or AB replac* N1 scheme*

S7 TI ((job search* or job applicat* or subsidi#ed Work)) or AB ((job search* or job* applicat* or subsidi#ed Work))

S8 TI ((help or assistance or incentive or initiative*or experience* or experiment or support or training or program* or scheme* or counsel* or course*)) or AB ((help or assistance or incentive or initiative*or experience* or experiment or support or training or program* or scheme* or counsel* or course*))

S9 s7 and s8

[7] The search strategy will have to be considerably modified for searching the IDEAS/Economist Online databases which do not allow complex searching. Even though these two databases contain similar references, we will search both in attempt to achieve as thorough a search as possible.

S10 TI ((job opportunity and basic skills program)) or AB ((job opportunity and basic skills program))

S11 TI ((job n1 finding or job-finding)) or AB ((job n1 finding or job-finding))

S12 TI ((help or assistance or program* or incentive or scheme* or training or counsel* or course*or initiative* or experience* or experiment* or support*)) or AB ((help or assistance or program* or incentive or scheme* or training or counsel* or course* or initiative* or experience* or experiment* or support*))

S13 S11 and S12

S14 (S1 or S2 or S5 or S6 or S9 or S10 or S13)

S15 DE "EMPLOYABILITY"

S16 (S14 or S15)

3.2.3 **Searching other resources**

Hand searching

Reference lists of included studies and reference lists of relevant reviews will be searched. The journals *Journal of Labor Economics* and *Labour Economics* will be hand searched for the years 2012 and available issues for 2013.

Grey literature

A general web search will be conducted using Google to identify potential unpublished studies. 'Advanced search' options will be used to refine the grey search strategy.

OpenSIGLE will be searched to identify potentially relevant reports from European grey literature (http://opensigle.inist.fr/).

The following Research Networks will be searched to identify working-papers and discussion-papers: IZA – Institute of the Study of Labor (www.iza.org), CEPR – Centre for Economic Policy Research (www.cepr.org), NBER – National Bureau of Economic Research (www.nber.org), CESifo (www.cesifo-group.de/portal/page/portal/ifoHome), and SSRN – Social Science Research Network (www.ssrn.com).

Copies of relevant documents from Internet-based sources will be made. We will record the exact URL and date of access.

Personal contacts

Personal contacts with national and international researchers will be made to identify unpublished reports and on-going studies.

3.3 DATA COLLECTION AND ANALYSIS

3.3.1 Selection of studies

Two review authors (ADK, TF) will independently read titles and available abstracts of reports and articles identified in the search to exclude those that are clearly irrelevant. Citations considered relevant by at least one review author will be retrieved in full text. If there is insufficient information in the title and abstract to assess eligibility, the full text will be retrieved. At least two review authors (TF, ADK) will read the full text versions to ascertain eligibility based on the selection criteria. In the first screening level (on the basis of title and abstract), a citation will only be moved to the second screening level if the answer is a yes or uncertain for the following criteria: (1) Are the participants unemployed individuals receiving unemployment insurance benefit during their unemployment? (2) Does the study focus on participation in ALMPs? (3) Is the report/article a quantitative evaluation study?

At the second level (on the basis of full text), eligibility inclusion criteria are extended to the following: (4) Does the study estimate an effect, using a control group or using an estimated counterfactual? (5) Does the study examine exits to employment?

The inclusion coding questions for level 1 and 2 will be piloted and adjusted if required (see Appendix 9.1). In the event of disagreement, a third review author and content specialist (GS) will be consulted. Reasons for excluding studies that otherwise might be expected to be eligible will be documented and presented in an appendix. The overall search and screening process will be illustrated in a flow-diagram. The review authors will not be blind to the names of the investigators, institutions, or the journals of publication of the articles.

3.3.2 Data extraction and management

At least two review authors (TF, ADK) will independently code and extract data from the included studies. Data and information will be extracted on: available characteristics of participants, intervention characteristics and control conditions,

research design, sample size, risk of bias and potential confounding factors, censoring, outcomes and results. A data extraction sheet will be piloted on several studies and revised as necessary (see Appendix 9.2). Extracted data will be stored electronically. Disagreements will be resolved by consulting an independent reviewer with extensive content and methods expertise (LP). Analysis will be conducted in RevMan5, SAS and STATA.

3.3.3 **Assessment of risk of bias in included studies**

We will assess the methodological quality of studies using a risk of bias model developed by Prof. Barnaby Reeves in association with the Cochrane Non-Randomised Studies Methods Group.[8] This model, an extension of the Cochrane Collaboration's risk of bias tool, covers risk of bias both in RCTs and in non-randomised studies that have a well-defined control group.

The extended model is organised as, and follows the same steps as, the risk of bias model described in the Cochrane Handbook, chapter 8 (Higgins & Green, 2011). The model is extended as follows:

1) The existing Cochrane risk of bias tool needs elaboration when assessing non-randomised studies because, for the latter, particular attention must be paid to selection bias and risk of confounding. The extended model therefore specifically incorporates a formalised and structured approach for the assessment of selection bias in non-randomised studies by adding an explicit item that focuses on confounding. This is based on a list of confounders considered important and defined in the protocol for the review. The assessment of confounding is made using a worksheet which is marked for each confounder according to whether it was considered by the researchers, the precision with which it was measured, the imbalance between groups, and the care with which adjustment was carried out (see Appendix 9.3). This assessment will inform the final risk of bias score for confounding.

2) RCTs must have a protocol that is defined prior to commencing recruitment, whereas non-randomised studies need not. This makes NRCTs at greater risk of bias compared to RCTs. The item concerning selective reporting therefore also requires assessment of the extent to which analyses (and potentially other choices) could have been manipulated to bias the findings reported (for example, by the choice of method of model fitting, and by the potential confounders that are considered). In addition, the model includes two separate yes/no items asking review authors

[8] This risk of bias model was introduced by Prof. Reeves at a workshop on risk of bias in non-randomised studies at SFI Campbell, February 2011. The model is a further development of work carried out in the Cochrane Non-Randomised Studies Method Group (NRSMG).

whether they judge the researchers to have had a pre-specified protocol and analysis plan.

3) Finally, the risk of bias assessment is refined, making it possible to discriminate between studies with varying degrees of risk. This refinement is achieved by the use of a 5-point scale for certain items (see the following section *Risk of bias judgement items* for details).

The refined assessment is pertinent when considering data synthesis as it operationalizes the identification of those studies with a very high risk of bias (especially in relation to non-randomised studies). The refinement increases transparency in assessment judgements and provides justification for excluding a study with a very high risk of bias from the meta-analysis.

Risk of bias judgement items

The risk of bias model used in this review is based on 9 items (see Appendix 9.3).

The 9 items refer to:

- **sequence generation** (Judged on a low/high risk/unclear scale)
- **allocation concealment** (Judged on a low/high risk/unclear scale)
- **confounders** (Judged on a 5 point scale/unclear)
- **blinding** (Judged on a 5 point scale/unclear)
- **incomplete outcome data** (Judged on a 5 point scale/unclear)
- **selective outcome reporting** (Judged on a 5 point scale/unclear)
- **other potential threats to validity** (Judged on a 5 point scale/unclear)
- **a priori protocol** (Judged on a yes/no/unclear scale)
- **a priori analysis plan** (Judged on a yes/no/unclear scale)

Confounding

An important part of the risk of bias assessment of nonrandomised studies (NRCT and NRS) is consideration of how the studies deal with confounding factors (see Appendix 9.3). Selection bias is understood as systematic baseline differences between groups which can therefore compromise comparability between groups. Baseline differences can be observable to the researcher (e.g. age and gender) and

unobservable (e.g. motivation and 'ability'). There is no single non-randomised study design that always solves the selection problem. Different designs attempt to provide solutions to the problem of potential selection bias under different assumptions, and consequently require different types of data. Designs particularly vary with respect to how they deal with selection on 'unobservable' factors. The "right" method depends on the model generating participation, i.e. assumptions about the nature of the process by which participants are selected into a programme.

As there is no universal correct way to construct counterfactuals, we will assess the extent to which the identifying assumptions (the assumption that makes it possible to identify the counterfactual) are explained and discussed. Preferably, study investigators should have made an effort to justify their choice of method.

Pre-specified list of confounders

For this review, we have identified the following observable confounding factors to be the most relevant: age, gender, education, ethnicity, labour market conditions, censoring and unemployment duration. In each study, we will assess whether these factors have been considered, and we will in addition assess other factors likely to be a source of confounding within the individual included studies. Furthermore, we will assess how each study deals with 'unobservables'.

The motivation for focusing on age, gender, education and ethnicity is that these are the major determinants of the risk of being unemployed (Layard, Nickell & Jackman, 2005).

Concerning unemployment duration, most studies find that the genuine duration dependence is negative, so that the longer the unemployment spell, the smaller the individual's chance of finding a job[9] (see Serneels, 2002, for an overview). Thus if the study does not control for unemployment duration, the effect of ALMP participation will be biased.

Another potential source of bias arises from differences in labour market conditions. If the study explores changes in ALMP participation over time or space as their source of variation, for example, it is very important to control for changes in labour market conditions over time (as a consequence of the business cycle, for example) or over space as the exit rate to employment most certainly will depend on this factor.

[9] The reason for this is that unemployment implies a loss of skills or that long periods of unemployment lead to a loss of self-confidence.

Censoring may also introduce bias. The effect of ALMP participation is often measured using survival data. Participants who do not leave the unemployment system before the end of the study are censored from the outcome data and have the potential for introducing bias if not adequately accounted for. Censoring of participants is therefore a potential threat, both in relation to the level of censoring and in relation to whether censoring is taken into account.

Assessment

Review authors (at least two, TF & ADK) will independently assess the risk of bias for each included study. Disagreements will be resolved by a third review author who has content and statistical expertise (GS). We will report the risk of bias assessment in risk of bias tables for each included study.

3.3.4 **Measures of treatment effect**

We expect that the treatment effect will be measured either as the impact on the hazard rate or as the impact on the probability of employment at some date or time interval after the completion of the programme.

Our main interest is to include studies in a meta-analysis where hazard ratios and variance are either reported or are calculable from the available data. The hazard rate is defined as the event rate at time *t* conditional on survival until time *t* or later. The effect is the difference in hazard rates between persons who have participated and persons who have not participated in ALMP (or the hazard rate they *would* have had *if* they had not participated in ALMP). Depending on the model specification, the effect may be given as a relative change in hazard rates or as an absolute change in hazard rates (van den Berg, 2001; Jenkins, 2005; Abbring & van den Berg, 2003). The effect size will be measured as log hazard ratios (HRs). The log hazard ratio measures the risk of event, in this case the risk of work, in the treatment group in comparison to the control group. Therefore, the individual included studies will be pooled using the log hazard ratio and variance. We will report the 95% confidence intervals. If log hazard ratios and variances are not reported then, dependent on available data, log hazard ratios and variances will be computed (Parmar et al., 1998; Lancaster, 1990). If studies report duration until employment the log hazard ratio and variance will be calculated if possible; depending on the model specification (Lancaster, 1990). Log hazard ratios and variances will be computed directly using the observed number of events and log rank expected number of

events if available; or indirectly if the p-value for the log-rank, Mantel-Haenszel or chi-squared test if one of these is reported (Sutton, Abrams, Jones, Sheldon & Song, 2000). Failing this individual participant data will *not* be requested to calculate log hazard ratios as this may introduce bias due to the time span of studies (the time span between the earliest we know of and the latest is 30 years).

The acceptable outcome measurement frequency for calculating hazard ratios in this review is three months or less. Where effect sizes are measured as log hazard ratios, studies reporting only outcomes measured on time intervals of more than three months will not be included in the meta-analysis.

We will pool separately studies reporting outcomes measured as 'duration until employment' or employment status at some date or time interval after completion of the programme where calculation of hazard ratios is not possible. For continuous outcomes (for example, weeks until employment) effect sizes will be calculated if standard deviations are available. Hedges' *g* will be used for estimating standardized mean differences (SMDs) and 95% confidence intervals will be reported. For employment status at some date or time interval after completion of the programme, we will use the relative risk ratio and variance, and will report the 95% confidence intervals. Outcomes measured at a single time point are inferior to log hazard ratios because the latter takes into account the full observation period, while the former does not (van den Berg, 2001; Jenkins, 2005; Abbring & van den Berg, 2003). In addition, the timing of point measures might be biased because researchers could have chosen to report results at time points with significant effects and not to report results at other time points where no difference between groups is apparent (selective reporting).

For secondary outcomes, duration of re-employment may be measured as hazard rates, in which case the effect size will be calculated as log hazard ratios with 95% confidence intervals. Alternatively, it may be reported directly as mean duration. The re-employment wage may be measured as the mean income at different time points or during different time periods. For such continuous outcomes, effect sizes will be calculated if standard deviations are available. Hedges' *g* will be used for estimating standardized mean differences (SMD) where scales measure the same outcomes in different ways. We will report the 95% confidence intervals.

3.3.5 **Unit of analysis issues**

We will take into account the unit of analysis of the studies to determine whether individuals were randomised in groups (i.e. cluster randomised trials), whether individuals may have undergone multiple interventions, whether there were multiple treatment groups, and whether several studies are based on the same data source, as follows.

3.3.5.1 Cluster randomisation

In cluster randomisation, statistical analysis errors can occur when the unit of allocation is different from the unit of analysis. In cluster-randomised trials, the elements are groups of individuals (e.g. region, PES office), rather than individuals themselves. In such studies, it is important to avoid biased standard errors if the unit-of-analysis is the individual. When an appropriate cluster analysis has been used (e.g. cluster summary statistics, robust standard errors, the use of the design effect to adjust standard errors, multilevel models and mixture models), effect estimates and their standard errors will be meta-analysed (Higgins & Green, 2011). In cases where study investigators have not applied appropriate analysis methods to control for clustering, we will attempt to estimate the intra-cluster correlation to correct standard errors.

The effective sample size of a single intervention group in a cluster-randomised trial is its original sample size divided by a quantity called the *design effect*. A common design effect is usually assumed across intervention groups. The design effect is 1+(m - 1)r, where m is the average cluster size and r is the *intra cluster correlation coefficient* (ICC). The standard errors of the effect estimates (from an analysis ignoring clustering) should be multiplied by the square root of the design effect. The total variance in the outcome can be partitioned into variance between groups (VBG) and variance within groups (VWG).The intra cluster correlation is calculated as VBG/(VBG+VWG). The ICC is seldom reported in the primary studies but if enough information is available we will approximate the intra-cluster correlation (see Donner et al., 2001) and correct standard errors.

3.3.5.2 Multiple interventions groups and multiple interventions per individuals

Multiple intervention groups (with different individuals) within a study with one control group will be pooled if appropriate and compared to the one control group. Multiple controls groups will only be pooled if appropriate.

A synthetic (the average) effect size will be calculated and used to avoid dependence problems. This method provides an unbiased estimate of the mean effect size parameter but overestimates the standard error. The use of random effects models when synthetic effect sizes are involved actually perform better in terms of standard errors than do fixed effects models (Hedges, 2007). However, tests of heterogeneity when synthetic effect sizes are included are rejected less often than nominal.

If pooling is not appropriate (when, for example, multiple interventions and/or control groups include the same individuals), only one intervention group will be coded and compared to the control group to avoid overlapping samples. Data from studies comparing different types of interventions/comparisons will be coded and analysed separately.

3.3.5.3 Multiple studies using the same sample of data

In some cases, several studies may have used the same sample of data (e.g. studies using the same register data). We will review all such studies, but will only include in the meta-analysis one estimate of the participation effect from each sample of data to avoid dependencies between the estimates of the participation effect. The choice of which estimate to include will be based on our quality assessment of the studies. We will choose the estimate from the study that we judge to have the least risk of bias, and we will primarily consider selection bias.

3.3.5.4 Multiple time points

When the results are measured at multiple time points, each outcome at each time point will be analysed in a separate meta-analysis with other comparable studies taking measures at a similar time point. As a guideline, these will be grouped together as follows: short-term (0 to <6 months after participation), medium term (6 to <12 months after participation) long term (at least 12 months after participation). However, should the studies provide viable reasons for an alternative choice of relevant and meaningful duration intervals for the analysis of outcomes, we will adjust the grouping.

3.3.6 **Dealing with missing data and incomplete data**

Missing data, attrition rates and censoring will be assessed in the included studies. Studies of ALMP typically estimate effects using data collected from administrative registers or by questionnaires. Studies using data collected from questionnaires are

subject to missing data. For studies using questionnaire data, a sensitivity analysis will be used to assess potential bias, and the extent to which the results might be biased by missing data will also be discussed.

The review authors will record attrition rates and reasons for attrition from included studies where possible. For studies in which the censoring level is high (more than 25%) or is not reported, a sensitivity analysis will be used to assess potential bias in the analysis. The extent to which the results might be biased by a high censoring level will be discussed.

Information on whether intention to treat analysis (ITT) was conducted (and if so, with what model) will also be recorded for RCTs and QRCTs. Sensitivity analysis will be performed to examine the impact of excluding trials in which adequate ITT analysis was not used.

3.3.7 **Assessment of heterogeneity**

Statistically significant heterogeneity among primary outcome studies will be tested with the Chi-squared (Q) test and measured with the I-squared statistic (Higgins, Thompson, Deeks & Altman, 2003). A significant Q ($P<.05$) and an I-squared value of at least 50% will be considered as indicators of heterogeneity. The sources of heterogeneity we will investigate are as follows: type of ALMP, type of participants, and labour market condition. The method is discussed in section 3.4.1.

3.3.8 **Assessment of reporting bias**

Reporting bias refers to both publication bias and selective reporting of outcome data and results. Bias from selective reporting of outcome data and results is one of the main items in the risk of bias model discussed in section 3.3.3, and is not addressed further here.

We will use funnel plots for information about possible publication bias if we find a sufficient number of studies (Higgins & Green, 2011). However, asymmetric funnel plots are not necessarily caused by publication bias (and publication bias does not necessarily cause asymmetry in a funnel plot). If asymmetry is present, we will consider possible reasons for this.

3.3.9 **Grading of evidence**

The quality of evidence will be assessed according to a systematic and explicit method (Guyatt et al., 2008). In order to indicate the extent to which one can be confident that an estimate of effect is correct, judgments about the quality of evidence will be made for each comparison and outcome. These judgments consider study design (RCTs, QRCTs, NRCTs and NRSs), study quality (detailed study design and execution), consistency of results (similarity of estimates of effect across studies) and directness (the extent to which people, interventions and outcome measures are similar to those of interest). The following definitions will be used in grading the quality of evidence for each outcome: **High**: further research is very unlikely to change our confidence in the estimate of effect. **Moderate**: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate. **Low**: further research is very likely to have an important impact on our confidence in the estimate of effect and may change the estimate. **Very low**: any estimate of effect is very uncertain.

3.4 DATA SYNTHESIS

Studies that have been coded with a very high risk of bias (scored 5 on the risk of bias scale) will not be included in the data synthesis

All follow-up durations reported in the primary studies will be recorded and we will conduct separate analyses for short-, medium- and long-term outcomes. Otherwise, included studies will be pooled where appropriate, dependent on heterogeneity in relation to covariates such as participant characteristics (age, gender, education and ethnicity), labour market conditions and unemployment duration.

As ALMPs vary in their content and deal with diverse populations of participants and labour market conditions, we will use a random effects model to estimate the overall effect. For subsequent analyses of moderator variables that may contribute to systematic variations, we will use the mixed-effects regression model. This model is appropriate if a predictor explaining some between-studies variation is available but there is a need to account for the remaining uncertainty (Hedges & Pigott, 2004; Konstantopoulos, 2006).

Lock-in effects

Lock-in effects may be considerable, depending on duration and intensity of programmes. Some studies will report effects net of locking-in effects and others will only report post programme effects. We will pool effects net of lock-in and post programme effects separately.

Counterfactual situation and threat effects

If ALMPs cover the entire labour force, the prospect of compulsory programme participation may result in an incentive effect, providing an incentive for unemployed persons to look for and return to work prior to programme participation (Jackman, 1994; Hansen & Tranæs, 1999). If a large proportion of those receiving unemployment insurance choose to leave unemployment (for example if they find a job or withdraw from the labour force) before they are required to participate in a compulsory programme, it may be very difficult to avoid selection bias in evaluations of the ALMP as discussed by Heckman, Ichimura and Todd (1997).

However, the incentive effects of ALMP have only been recognized recently (Geerdsen et al., 2011). We do not expect that many studies try to correct for this type of selection bias. To deal with the issue we will, where possible, distinguish between the counterfactual situations. The main distinction between counterfactual situations is whether the studies estimate an effect relative to a world without compulsory ALMP (such as when the control group is never going to participate and there is no threat effect) or they estimate an effect relative to a situation where compulsory participation will occur at a later point in time (a system of ALMPs and exposure to a threat effect).

3.4.1 Moderator analysis and investigation of heterogeneity

If heterogeneity is judged to be large, we will investigate the following factors with the aim of explaining observed heterogeneity: type of ALMP (training, private sector incentive programmes, direct employment programmes in the public sector and job search assistance programmes); study-level summaries of participant characteristics (e.g. studies considering a specific age group, gender or educational level or studies where separate effects for men/women, young/old or low/high educational level are available); and labour market conditions.

If the number of included studies is sufficient (at least 10 degrees of freedom) and there is sufficient variation in the covariates, we will perform moderator analyses

(multiple meta-regression using the mixed model) to explore how observed variables are related to heterogeneity. We will estimate the (new) residual variance component to be used in a weighted least-squares analysis conditional on this variance component estimate. The residual variance component will be estimated using the method-of-moments estimator (Hartung, Knapp & Sinha, 2008; Konstantopoulos, 2006). We will report the 95% confidence intervals for regression parameters. We will estimate the correlations between the covariates and consider the possibility of confounding. Conclusions from meta-regression analysis will be drawn with caution and will not be based on significance tests.

Otherwise, single factor subgroup analysis will be performed. The assessment of any difference between subgroups will be based on 95% confidence intervals. No conclusions from subgroup analyses will be drawn, and interpretation of relationships will be cautious as they are based on subdivision of studies and indirect comparisons.

In general, the strength of inference regarding differences in treatment effects among subgroups is controversial. However, making inferences about different effect sizes among subgroups on the basis of between-study differences entails a higher risk compared to inferences made on the basis of within-study differences (Oxman & Guyatt, 1992). We will therefore use within-study differences were possible.

We will also consider the degree of consistency of differences as making inferences about different effect sizes among subgroups entails a higher risk when the difference is not consistent within the studies (Oxman & Guyatt, 1992).

3.4.2 **Sensitivity analysis**

Sensitivity analysis will be carried out by restricting the meta-analysis to a subset of the totality of studies included in the original meta-analysis. Sensitivity analysis will be used to evaluate whether the pooled effect sizes are robust across study designs and components of methodological quality. For methodological quality, we will consider sensitivity analysis for each major component of the risk of bias checklists and will restrict the analysis to studies with a low risk of bias.

Sensitivity analysis will further be used to examine the strength of conclusions in relation to the quality of data, such as where outcome measures are based on

different time intervals and whether the data is derived from questionnaires or administrative registers.

4 Sources of support

4.1 INTERNAL SOURCES

SFI Campbell, The Danish National Centre for Social Research.

4.2 EXTERNAL SOURCES

None

5 Contributions of authors

Filges, Knudsen, Smedslund and Geerdsen contributed to the writing and revising of this protocol. The search strategy was developed by Filges and Jørgensen.

6 Acknowledgements

We appreciate the guidance and comments of William Turner, Nick Huband, Emily E. Tanner-Smith, Karianne T. Hammerstrøm and two anonymous peer reviewers.

7 Potential conflicts of interest

The authors have no vested interest in the outcomes of this review, nor any incentive to represent findings in a biased manner.

8 References

Abbring, J.H. & van den Berg, G.J. (2003).The nonparametric identification of treatment effects in duration models, *Econometrica, 71*(5), 1491-1517.

Black, D.A., J. A. Smith, M. C. Berger, & J. N. Brett (2003). Is the threat of reemployment services more effective than the services themselves? Evidence from random assignment in the UI system. *American Economic Journal, 93*(4), 1313-1327.

Calmfors, L. (1994). 'Active labour market policy and unemployment – a framework for the analysis of crucial design features', *OECD Economic Studies 22*. Available online:
http://www.oecd.org/economy/productivityandlongtermgrowth/33936463.pdf

Calmfors, L. (1995). What can we expect from active labour market policy? *Konjunkturpolitik 43*, 1-30.

Card D., Kluve, J. & Weber, A. (2010). Active labor market policy evaluations: A meta-analysis. *The Economic Journal, 120*(548), 452–477.

Congressional Budget Office (CBO) (2010). Social security disability insurance: Participation trends and their fiscal implications. *Policy paper. CBO. Available online:*
http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/116xx/doc11673/07-22-ssdisabilityins_brief.pdf. CBO: Washington, DC. July 22, 2010.

Donner, A., Piaggio, G. & Villar, J. (2001). Statistical methods for the meta-analysis of cluster randomized trials. *Statistical Methods in Medical Research 2001, 10*(5), 325-38.

Eurostat (2005). European Social Statistics: Labour market policy – expenditure and participants - Data 2003, Office for Official Publications of the European Commission, Luxembourg.

Filges, T., Kennes, J., Larsen, B. & Tranæs, T. (2011), Labour market programmes and the equity-efficiency trade-off. *Journal of Macroeconomics, 33*(4), 738-53.

Geerdsen, L. P. (2003). Marginalisation processes in the Danish labour market, Ph.D. thesis, *The Danish National Institute of Social Research, Report 03:24. Available online:* http://www.sfi.dk/kopi_af_resultater-4724.aspx?Action=1&NewsId=283&PID=9404

Geerdsen, L.P., Bjørn, N.H., Filges, T. & Jensen, P. (2011). The threat of active labour market programmes for unemployed. Campbell Collaboration, Review submitted.

Gerfin, M. & Lechner, M. (2002). A microeconometric evaluation of the active labour market policy in Switzerland. *The Economic Journal, 112*(482), 854-893.

Guyatt, G.H., Oxman, A.D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., Schünemann HJ; GRADE Working Group.(2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal, 336*, 924-6.

Hansen, C. T. & T. Tranæs (1999*)*. Optimal workfare in a society of workers and nonworkers. In Andersen, T. M, Jensen S. E. H., and Risager, O. (Eds). Macroeconomic perspectives on the Danish economy. London: MacMillan Press, pp. 335-358.

Hartung, J., Knapp, G. & Sinha, B. K. (2008). *Statistical meta-analysis with applications*. New York: Wiley.

Heckman, J. J., Ichimura, H. & Todd, P. E. (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training programme, *Review of Economic Studies, 64*(4), 605-54.

Heckman, J. J, Lalonde, R. J. & Smith, J. A (1999). The economics and econometrics of active labour market programs. In Ashenfelter, O. and Card, D. (Eds.). *Handbook of Labor Economics 3*, pp.1865-2097. Amsterdam: Elsevier.

Hedges, L. V. (2007). Meta-analysis. In: Rao, C.R. (Eds.). *The Handbook of Statistics*, pp.919-53. Amsterdam: Elsevier.

Hedges, L.W. & Pigott, T.D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9*(4), 426–445.

Higgins, J.P.T. & Green, S. (eds) (2011). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011].Wiley-Blackwell The Cochrane Collaboration. Available from www.cochrane-handbook.org.

Higgins, J.P., Thompson, S.G., Deeks J.J. & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557-60.

Høgelund, J. & Holm, A. (2005), Returning the long-term sick-listed to work: The effects of educational measures and employer separations in Denmark.In: Saunders, P. (Eds.). *Welfare to Work in Practice. Social Security and Participation in Economic and Social Life*. International Studies on Social Security, Vol. 10, Aldershot: Ashgate.

International Labour Office (1990), *Employment, unemployment and underemployment. An ILO Manual on Concepts and Methods*, Geneva, International Labour Office.

Jackman, R. (1994). What can active labour market policy do? Swedish Economic Policy Review, 1, 221-257.

Jenkins, Stephen P. (2005), Survival analysis, Lecture notes, 18 July 2005. http://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf

Kluve, J. (2006). The effectiveness of European active labor market policy. *IZA DP No. 2018. Available online: http://ftp.iza.org/dp2018.pdf*

Kluve, J., Lehmann H. & Schmidt, C.M. (1999). Active labor market policies in Poland: Human capital enhancement, stigmatization or benefit churning. *Journal of Comparative Economics 27*(1), 61-89.

Kluve, J. & Schmidt, C.M. (2002). Can training and employment subsidies combat European unemployment? *Economic Policy, 17*(35), 409-448.

Kluve, J., Card, D., Fertig, M., Góra, M., Jacobi, L., Jensen, P. et al. (2007). *Active labor market policies in Europe: Performance and perspectives*. Essen: Springer.

Konstantopoulos, S. (2006). Fixed and mixed effects models in meta-analysis. *IZA DP no. 2198*.

Lancaster, T. (1990). The econometric analysis of transition data. *Econometric Society Monographs No. 17*, Cambridge: Cambridge University Press.

Layard, Richard, Nickell, S. & Jackman, R. (2005). *Unemployment*. Oxford University Press.

Martin, J.P. (2000). What works among active labour market policies: Evidence from OECD countries' experiences. *OECD Economic Studies, No. 30*(2000/1), 79-113.

Martin. J. P. & Grubb D. (2001). What works and for whom: a review of OECD countries' experiences with active labour market policies, *Swedish Economic Policy Review, 8*(2), 9-56.

Mortensen, D. T (1987): Job search and labor market analysis. In Ashenfelter, O. & Layard, R. (Eds.). *Handbook of Labor Economics*, vol. 2, North-Holland, chapter 15, pp. 849–919.

OECD (2004), *Employment Outlook*, OECD: Paris.

OECD (2007), *Benefits and Wages*, OECD Indicators.

Oxman, A. & G.H. Guyatt, A. (1992) Consumer's Guide to Subgroup Analyses, *Annals of Internal Medicine, 116*(1), 78-84.

Parmar, M.K.B., Torri, V. & Stewart, L. (1998). Extracting summary statistics to perform meta analyses of the published literature for survival endpoints. *Statistics in Medicine, 17*, 2815-2834.

Pissarides, C. A. (2000). *Equilibrium unemployment theory*, 2nd ed., Cambridge (MA): MIT Press.

Serneels, P. (2002). Explaining non-negative duration dependence among the unemployed, *CSAE Working Paper Series 2002-13*, Centre for the Study of African Economies, University of Oxford.

Smedslund G., Hagen, K.B., Steiro, A., Johme, T., Dalsbø, T.K. & Rud, M.G. (2006). Work programmes for welfare recipients, *Campbell Collaboration Review*.

Smith, J.A. (2000a). A critical survey of empirical methods for evaluating active labor market policies. *Swiss Journal of Economics and Statistics, 136*(3), 247-268.

Smith, J.A. (2000b). Evaluating active labor market policies: Lessons from North America. *Working Paper, August 15*.

Statistics Denmark, 2009 (in Danish). The registered unemployment statistics, Labour Force Survey (AKU). (Den registrerede ledighedsstatistik Arbejdskraftundersøgelsen (AKU)), *Danmarks Statistik, 20. maj*.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods of meta-analysis in medical research*. New York: Wiley.

Van den Berg, G.J., (2001). Duration models: Specification, identification, and multiple durations. In Heckman, J.J. & Leamer, E. (Eds*)*, *Handbook of Econometrics* (chapter 55), Volume V (North-Holland, Amsterdam), 2001.

9 Appendices

9.1 FIRST AND SECOND LEVEL SCREENING

First level screening is on the basis of titles and abstracts. Second level is on the basis of full text

Reference id. No. :
Study id. No.:
Reviewers initials:
Source:
Year of publication:
Duration of study:
Country/countries of origin
Author

The study will be excluded if one or more of the answers to question 1-3 are 'No'. If the answers to question 1 to 3 are 'Yes' or 'Uncertain', then the full text of the study will be retrieved for second level eligibility. All unanswered questions need to be posed again on the basis of the full text. If not enough information is available, or if the study is unclear, the author of the study will be contacted if possible.

First level screening questions are based on titles and abstracts

1. Are the participants' unemployed individuals receiving unemployment insurance benefits during their unemployment?
   Yes - include
   No – if no then stop here and exclude
   Uncertain - include

Question 1 guidance:

This includes only unemployment insurance (UI) benefits. We are not interested in other types of unemployment benefits such as unemployment assistance benefits and social assistance benefits. In most OECD countries a secondary benefit is available for those who have exhausted regular unemployment insurance benefits. These are known as 'social assistance benefits'. Unlike UI benefits, social assistance benefits are generally means-tested, pay a lower level of benefit, and have no time limit.

2. Does the study focus on ALMP (Active Labour Market Programme) participation?
   Yes - include
   No – if no then stop here and exclude
   Uncertain - include

Question 2 guidance:
The intervention is ALMP participation. The programme may be compulsory (participants have to participate to continue receiving unemployment insurance benefits) or voluntary. This intervention can be referred to in different ways, e.g. job search assistance, employment counselling, job preparation activities, education and training and private or public sector employment programmes.

3. Is this study a primary quantitative study?
   Yes - include
   No – if no then stop here and exclude
   Uncertain - include

Question 3 guidance:
We are only interested in primary quantitative studies, where the authors have analyzed the data. We are not interested in theoretical papers on the topic or surveys/reviews of studies of the topic. (This question may be difficult to answer on the base of titles and abstracts alone.)

Second level screening questions based on full text

4. Does the study estimate an effect, using a control group or using an estimated counterfactual?
   Yes - include
   No – if no then stop here and exclude
   Uncertain - include

Question 4 guidance
E.g. 1) Randomised controlled trials including cluster randomisation and quasi randomised controlled study designs (i.e. participants are allocated by means such

as alternate allocation, person's birth date, the date of the week or month, case number or alphabetical order), 2) non randomised controlled study designs (i.e. quasi-experimental designs) such as controlled two group study designs or 3) study designs based on observational data, where the effect is estimated by statistical methods.

5. Does the study examine exits to employment?
   Yes – include
   No – if no then stop here and exclude
   Uncertain – include

Question 3 guidance:

The primary outcome is exits to employment. Studies only looking at exits to other destinations (such as other kinds of benefits or out of the labour force) will not be included. Studies who do not distinguish between destinations will be included.

9.2 DATA EXTRACTION

Names of author(s)
Title
Language
Journal
Year
Country
Target group (age, gender, education, eligibility requirements for benefits)
Duration of benefit period prior to ALMP participation
Is the programme compulsory?
How are individuals informed about ALMP?
Types of ALMPs (labour market training/education, private sector programmes, public sector employment programmes, job search assistance)
Do individuals attend more than one programme?
Benefit level during ALMP participation (more/less than non-participation)
Duration of a ALMP (days, weeks, months)

How many hours a week/month do individuals participate in ALMP?
Sanctions if individual refuses to participate in a programme
Labour market conditions (unemployment rate, vacancy rate, labour market tightness, etc.)
Type of data used in study (administrative, questionnaire, other (specify))
Time period covered by analysis
Time interval the outcome measure is based on (daily, weekly, monthly, etc.)
Which counterfactual situation is participation compared to? (the control group is never going to participate, participation will occur at a later point in time
Is the measured effect net of lock-in effects
Sample size

Outcome measures

Instructions: Please enter outcome measures in the order in which they are described in the report. Note that a single outcome measure can be completed by multiple sources and at multiple points in time (data from specific sources and time-points will be entered later).

#	Outcome & measure	Reliability & Validity	Format	Direction	Pg# & notes
1		Info from: Other samples This sample Unclear Info provided:	Dichotomy Continuous	High score or event is Positive Negative Can't tell	

* Repeat as needed

OUT COME DATA

DICHOTOMOUS OUTCOME DATA

OUTCOME	TIME POINT (s) (record exact time from participation, there may be more than one, record them all)	SOURCE	VALID Ns	CASES	NON-CASES	STATISTICS	Pg. # & NOTES
		Questionnaire Admin data Other (specify) Unclear	Participation Comparison 	Participation Comparison 	Participation Comparison 	RR (risk ratio) OR (odds ratio) SE (standard error) 95% CI DF P- value (enter exact p value if available) Chi2 Other	

Repeat as needed

OUT COME DATA

TIME-TO-EVENT OUTCOME DATA

OUTCOME	TIME POINT (s) (record exact time from participation, there may be more than one, record them all)	SOURCE	STATISTICS	Pg. # & NOTES
		Questionnaire Admin data Other (specify) Unclear	HR (hazard ratio) SE (standard error) 95% CI DF P- value (enter exact p value if available) Chi2 Other	

Repeat as needed

CONTINUOUS OUTCOME DATA

OUTCOME	TIME POINT (s) (record exact time from participation, there may be more than one, record them all)	SOURCE (specify)	VALID Ns	Means	SDs	STATISTICS	Pg. # & NOTES
		Questionnaire Admin data Other (specify) Unclear	Participation	Participation	Participation	P t F Df ES Other	
			Comparison	Comparison	Comparison		

*Repeat as need

9.3 ASSESSMENT OF RISK OF BIAS IN INCLUDED STUDIES

Risk of bias table

Item	**Judgement**[a]	**Description** (quote from paper, or describe key information)
1. Sequence generation		
2. Allocation concealment		
3. Confounding[b,c]		
4. Blinding?[b]		
5. Incomplete outcome data addressed?[b]		
6. Free of selective reporting?[b]		
7. Free of other bias?		
8. A priori protocol?[d]		
9. A priori analysis plan?[e]		

[a] Some items on low/high risk/unclear scale (double-line border), some on 5 point scale/unclear (single line border), some on yes/no/unclear scale (dashed border). For all items, record "unclear" if inadequate reporting prevents a judgement being made.

[b] For each outcome in the study.

[c] This item is only used for NRCTs and NRSs. It is based on list of confounders considered important at the outset and defined in the protocol for the review (*assessment against worksheet*).

[d] Did the researchers write a protocol defining the study population, intervention and comparator, primary and other outcomes, data collection methods, etc. in advance of starting the study?

[e] Did the researchers have an analysis plan defining the primary and other outcomes, statistical methods, subgroup analyses, etc. in advance of starting the study?

Risk of bias tool

Studies for which RoB tool is intended

The risk of bias model was developed by Prof. Barnaby Reeves in association with the Cochrane Non-Randomised Studies Methods Group.[10] This model, an extension of the Cochrane Collaboration's risk of bias tool, covers risk of bias in both randomised controlled trials (RCTs and QRCTs) and in non-randomised studies (NRCTs and NRSs).

The point of departure for the risk of bias model is the Cochrane Handbook for Systematic Reviews of interventions (Higgins & Green, 2011). The existing Cochrane risk of bias tool needs elaboration when assessing non-randomised studies because, for non-randomised studies, particular attention should be paid to selection bias / risk of confounding. Additional item on confounding is used only for non-randomised studies (NRCTs and NRSs) and is not used for randomised controlled trials (RCTs and QRCTs).

Assessment of risk of bias

Issues when using modified RoB tool to assess included non-randomised studies:

- Use existing principle: score judgment and provide information (preferably direct quote) to support judgment
- Additional item on confounding used only for non-randomised studies (NRCTs and NRSs).
- 5-point scale for some items (distinguish "unclear" from intermediate risk of bias).
- Keep in mind the general philosophy – assessment is not about whether researchers could have done better but about risk of bias; the assessment tool must be used in a standard way whatever the difficulty / circumstances of investigating the research question of interest and whatever the study design used.
- Anchors: "1/No/low risk" of bias should correspond to a high quality RCT. "5/high risk" of bias should correspond to a risk of bias that means the findings should not be considered (too risky, too much bias, more likely to mislead than inform)

1. Sequence generation
   - Low/high/unclear RoB item
   - Always high RoB (not random) for a non-randomised study
   - Might argue that this item redundant for NRS since always high – but important to include in RoB table ('level playing field' argument)

2. Allocation concealment
   - Low/high/unclear RoB item
   - Potentially low RoB for a non-randomised study, e.g. quasi-randomised (so high RoB to sequence generation) but concealed (reviewer judges that the people making decisions about including participants didn't know how allocation was being done, e.g. odd/even date of birth/hospital number)

3. RoB from confounding (additional item for NRCT and NRS; assess for each outcome)
   - Assumes a pre-specified list of potential confounders defined in the protocol

[10] This risk of bias model was introduced by Prof. Reeves at a workshop on risk of bias in non-randomised studies at SFI Campbell, February 2011. The model is a further development of work carried out in the Cochrane Non-Randomised Studies Method Group (NRSMG).

- Low(1) / 2 / 3 / 4 / high(5) / unclear RoB item
- Judgment needs to factor in:
    - proportion of confounders (from pre-specified list) that were considered
    - whether most important confounders (from pre-specified list) were considered
    - resolution/precision with which confounders were measured
    - extent of imbalance between groups at baseline
    - care with which adjustment was done (typically a judgment about the statistical modeling carried out by authors)
- Low RoB requires that all important confounders are balanced at baseline (not primarily/not only a statistical judgment OR measured 'well' and 'carefully' controlled for in the analysis.

Assess against pre-specified worksheet. Reviewers will make a RoB judgment about each factor first and then 'eyeball' these for the judgment RoB table.

4. RoB from lack of blinding (assess for each outcome, as per existing RoB tool)
- Low(1) / 2 / 3 / 4 / high(5) / unclear RoB item
- Judgment needs to factor in:
    - nature of outcome (subjective / objective; source of information)
    - who was / was not blinded and the risk that those who were not blinded could introduce performance or detection bias
    - see Ch.8

5. RoB from incomplete outcome data (assess for each outcome, as per existing RoB tool)
- Low(1) / 2 / 3 / 4 / high(5) / unclear RoB item
- Judgment needs to factor in:
    - reasons for missing data
    - whether amount of missing data balanced across groups, with similar reasons
    - whether censoring is less than or equal to 25% and taken into account
    - see Ch.8

6. RoB from selective reporting (assess for each outcome, NB different to existing Ch.8 recommendation)
- Low(1) / 2 / 3 / 4 / high(5) /unclear RoB item
- Judgment needs to factor in:
    - existing RoB guidance on selective outcome reporting (see Ch.8)
    - also, extent to which analyses (and potentially other choices) could have been manipulated to bias the findings reported, e.g. choice of method of model fitting, potential confounders considered / included
    - look for evidence that there was a protocol in advance of doing any analysis / obtaining the data (difficult unless explicitly reported); NRS very different from RCTs. RCTs must have a protocol in advance of starting to recruit (for REC/IRB/other regulatory approval); NRS need not (especially older studies)
    - Hence, separate yes/no items asking reviewers whether they think the researchers had a pre-specified protocol and analysis plan.

7. RoB from other bias (assess for each outcome, NB different to existing Ch.8 recommendation)
- Low(1) / 2 / 3 / 4 / high(5) /unclear RoB item
- Judgment needs to factor in:
    - existing RoB guidance on other potential threats to validity (see Ch.8)

- also, assess whether suitable cluster analysis is used (e.g. cluster summary statistics, robust standard errors, the use of the design effect to adjust standard errors, multilevel models and mixture models), if assignment of units to treatment is clustered

Confounding Worksheet

Assessment of how researchers dealt with confounding	
Method for *identifying* relevant confounders described by researchers: yes no If yes, describe the method used:	☐ ☐
Relevant confounders described: yes no List confounders described on next page	☐ ☐
Method used for controlling for confounding At design stage (e.g. matching, regression discontinuity, instrument variable): .... At analysis stage (e.g. stratification, regression, difference-indifference): .... Describe confounders controlled for below	

Confounders described by researchers

Tick (yes[0]/no[1] judgment) if confounder considered by the researchers [Cons'd?]

Score (1[good precision] to 5[poor precision]) precision with which confounder measured

Score (1[balanced] to 5[major imbalance]) imbalance between groups

Score (1[very careful] to 5[not at all careful]) care with which adjustment for confounder was carried out

Confounder	Considered	Precision	Imbalance	Adjustment
Gender	☐	☐	☐	☐
Age	☐	☐	☐	☐
Ethnicity	☐	☐	☐	☐
Education	☐	☐	☐	☐
Labour market condition	☐	☐	☐	☐
Unemployment duration	☐	☐	☐	☐
Unobservables[11]	☐	Irrelevant	☐	☐
Censoring	☐	☐	☐	☐
Other:	☐	☐	☐	☐

[11] See user guide for unobservables

User guide for unobservables

Selection bias is understood as systematic baseline differences between groups and can therefore compromise comparability between groups. Baseline differences can be observable (e.g. age and gender) and unobservable (to the researcher; e.g. motivation and 'ability'). There is no single non-randomised study design that always solves the selection problem. Different designs solve the selection problem under different assumptions and require different types of data. Especially how different designs deal with selection on unobservables varies. The "right" method depends on the model generating participation, i.e. assumptions about the nature of the process by which participants are selected into a programme.

As there is no universal correct way to construct counterfactuals we will assess the extent to which the identifying assumptions (the assumption that makes it possible to identify the counterfactual) are explained and discussed (preferably the authors should make an effort to justify their choice of method). We will look for evidence that authors using e.g. (this is NOT an exhaustive list):

Natural experiments:
Discuss whether they face a truly random allocation of participants and that there is no change of behaviour in anticipation of e.g. policy rules.

Instrument variable (IV):
Explain and discuss the assumption that the instrument variable does not affect outcomes other than through their effect on participation.

Matching (including propensity scores):
Explain and discuss the assumption that there is no selection on unobservables, only selection on observables.

(Multivariate, multiple) Regression:
Explain and discuss the assumption that there is no selection on unobservables, only selection on observables. Further discuss the extent to which they compare comparable people.

Regression Discontinuity (RD):
Explain and discuss the assumption that there is a (strict!) RD treatment rule. It must not be changeable by the agent in an effort to obtain or avoid treatment. Continuity in the expected impact at the discontinuity is required.

Difference-in-difference (Treatment-control-before-after):
Explain and discuss the assumption that outcomes of participants and nonparticipants evolve over time in the same way.