



Reducing Unemployment Benefit Duration to Increase Job Finding Rates: Protocol for a Systematic Review

Trine Filges, Anders Bruun Jonassen and Anne-Marie Klint Jørgensen

Submitted to the Coordinating Group of:

<input type="checkbox"/>	Crime and Justice
<input type="checkbox"/>	Education
<input type="checkbox"/>	Disability
<input type="checkbox"/>	International Development
<input type="checkbox"/>	Nutrition
<input checked="" type="checkbox"/>	Social Welfare
<input type="checkbox"/>	Other:

Plans to co-register:

<input checked="" type="checkbox"/>	No		
<input type="checkbox"/>	Yes	<input type="checkbox"/> Cochrane	<input type="checkbox"/> Other
<input type="checkbox"/>	Maybe		

Date Submitted:

Date Revision Submitted:

Approval Date:

Publication Date: 01 November 2015

BACKGROUND

The Problem, Condition or Issue

Benefit programmes protect individuals against loss of income and provide unemployed individuals the possibility of finding a better match between their qualifications and job vacancies. This positive aspect of inducing workers to achieve better job matches has been shown to increase economic efficiency (Acemoglu & Shimer, 1999; Marimon & Zilibotti, 1999).

However, unemployment benefits may also distort incentives by subsidizing long and unproductive job searches. In fact, the generosity of unemployment benefits is generally considered the main factor by which benefit systems affect unemployment. From a societal point of view, therefore, the optimal unemployment benefit system will balance considerations for protection with those for distortion (Feldstein, 2005; Mortensen, 1987).

Theory suggests that putting a limit on benefit duration will tend to accelerate job search from the beginning of the unemployment spell and thereby shorten unemployment duration (Pissarides, 2000). Thus, generosity of benefits is determined not only by the amount paid but also by the duration of benefit entitlement. In the US, replacement rates¹ are low and duration is short compared to benefit systems in most European countries. In 2005 the maximum duration of unemployment insurance entitlement among OECD countries² was shortest in the US at 6 months³ and longest in Denmark, Norway, Portugal, the Netherlands, France, Finland and Spain, varying between 23 and 48 months (OECD, 2007). At the same time, the gross initial replacement rate was around 50% in the US, while varying between 62% and 90% in the aforementioned European countries.

The lower level of generosity of benefits in the US compared to Europe is consistent with the observation of higher levels of active searches and a greater willingness to accept inferior jobs by unemployed workers in the US compared to Europe (Layard, Nickell & Jackman, 2005). As a consequence European policy-makers may be tempted to reduce the generosity of unemployment systems in order to reduce high unemployment levels⁴. While lowering the replacement rate may be politically

¹ The replacement rate is the ratio of the unemployment benefit to that of previous earnings.

² For a 40-year-old single worker without children and with a 22-year employment record.

³ The maximum duration was also around six months in the Czech Republic, the Slovak Republic, and the United Kingdom.

⁴ An alternative could be attaching behavioural conditions in terms of required job search and required acceptance of job offers to benefit receipt.

intractable (indeed, examples of reductions of benefit rates and amounts are rare), the length of the unemployment benefit entitlement period is often used as a political instrument to improve work incentives for the unemployed. In Spain, for example, the benefit period was altered in 1992, in Slovenia in 1998, in Norway in 1997, in the UK in 1996, in Denmark in 1996, 1998 and 1999, and, more recently, in the Czech Republic in 2004, in Hungary and Portugal in 2006, and in Denmark again in 2010.

The important public policy question is whether a more generous unemployment benefit system is causally related to higher unemployment rates. As pointed out in Card and Riddell (1993), there can be several complementary explanations for high unemployment rates, including differences in the fraction of nonworking time that is reported as unemployment (particularly among individuals with very low levels of labour supply), and differences in the overall distributions of working and nonworking time. Recent research on the effect of extended duration of unemployment insurance benefits in the US shows that benefit extensions raised the unemployment rate, but at least half of the effect is attributable to reduced labour force exit among the unemployed rather than to the changes in reemployment rates that are of greater policy concern (Rothstein, 2011).

This review will focus on the effect on job finding rates of reducing the maximum duration of entitlement of unemployment benefits, and secondarily on the effects on the quality of these jobs.

The Intervention

The intervention of interest is reduction in the maximum duration of entitlement of any kind of unemployment benefit with a known expiration date. The benefits may be unemployment insurance (UI) benefits, or they may be unemployment assistance (UA)/social assistance (SA) benefits as long as they have a known expiration date.

In the majority of OECD countries, the UI benefit has a time-limit. In fact, only Belgium has an unlimited UI period. In other countries, the maximum duration varies between 6 months (as for example in the UK and the US) and 36 months (in Iceland).

In most OECD countries, a secondary benefit is available for those who have exhausted regular UI benefits. This is known as SA benefits. Unlike UI benefits, SA benefits are generally means-tested without any necessary connection to past employment; they pay a lower level of benefit and are indefinite. We know of only one example of a SA benefit with a time limit: the Temporary Assistance to Needy

Families (TANF) which is available in the US. The federal government requires states to impose between 2- or 5-year limits on TANF (Gustafson & Levine, 1997). In a minority of OECD countries, UA benefits are paid after exhaustion of UI benefits. Like SA benefits, they are generally means-tested, pay a lower level of benefits and, excepting Hungary, Portugal and Sweden, are indefinite. Unemployment benefits with an indefinite time limit or non-financial benefits will be excluded from this review.

How the Intervention Might Work

Search theory offers an explanation for how this intervention might work. According to search theory, one can derive a relationship between the job-finding rate and the parameters of the benefit system, in particular the maximum benefit duration and the replacement rate (Mortensen, 1977). This relationship is driven by adjustments in search effort and reservation wages. The reservation wage is the minimum wage at which the unemployed are willing to accept a job. Forward-looking unemployed workers chose their current search effort and reservation wage in order to maximize the sum of the utility flow realized during the current period, plus the expected discounted future utility flow given that an optimal strategy will be pursued in every future period. The current search effort and reservation wage are thus affected by the future level of benefits. When the benefit period expires, the unemployed person experiences a potentially large drop in income. As the time of benefit exhaustion approaches, the value to that person of remaining unemployed falls, implying a higher search effort and/or a fall in the reservation wage, leading to a higher exit rate out of unemployment (Mortensen, 1977). This non-stationarity implies that unemployed individuals with different lengths of benefit entitlement have different optimal paths of reservation wage and search effort over time (van den Berg 1990).

A shorter entitlement period gives the unemployed individual a stronger incentive to quickly gain employment in order to avoid the drop in income after the exhaustion date. How strong the incentive is depends on the magnitude of the income drop. If no secondary benefit is available for those who have exhausted their current benefit, the incentive to gain employment will be stronger. If an increased job finding rate is mainly driven by lowering the reservation wage, a lower job match quality is to be expected, for example, in the form of lower wages and/or lower re-employment duration.

A number of factors may have an impact on the magnitude of the expected increase in the job finding rate. In general, the overall labour market conditions (i.e. the vacancy rate⁵ and, in particular, the unemployment rate) have an impact on the availability of and competition for jobs. If the vacancy rate is high (i.e. the number of vacancies is high in relation to job seekers) we would expect a bigger effect on job finding rates than if the vacancy rate is low. We would further expect a lower effect if the unemployment rate is high, regardless of the vacancy rate. If the vacancy rate is low coincident with a high unemployment rate, competition for available jobs is likely to be high. If the vacancy rate is high coincident with a high unemployment rate, it suggests mismatch in the labour market (i.e., the process by which vacant jobs and job seekers meet is not efficient) (Filges & Larsen, 2000; Pissarides, 2000).

Whether compulsory participation in active labour market programmes is part of the unemployment system may also have an impact on the effect of maximum duration of entitlement. The compulsory aspect of activation may provide an incentive for unemployed individuals to look for and return to work prior to programme participation; the so called threat effect. Filges and Hansen (2015) summarize the available evidence on the threat effect of active labour market programmes and report a significant threat effect of compulsory participation in active labour market programmes. Further, actual participation in active labour market programmes may improve some of the participants' qualifications, thus helping them to find a job. Alternatively, active labour market programmes may have negative stigmatization and signalling effects to employers. Programmes associated with participants having poor employment prospect may carry a stigma. Because of asymmetric information, employers do not 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 programmes as low productivity workers or workers with tenuous labour market attachment (Kluve et al. 1999; Kluve et al., 2007).

A recent systematic review by Filges et al. (2015) investigated the effect of participating in active labour market programmes and found that there is a significant positive effect, although small, of participating in active labour market programmes. The effect reported in Filges et al. (2015) is however a pure post-programme effect of active labour market programmes; it refers to the period after participation in a programme. The net effect of active labour market programme participation on job-finding rates is, however, composed of two separate effects: a

⁵ The number of unfilled jobs expressed as a proportion of the labour force.

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 combination of the two effects, lock-in and post-programme, consequently determines the net effects of active labour market programme participation on unemployment duration.

These additional effects on the search behaviour and employment prospects when compulsory participation in active labour market programmes is part of the unemployment system may dampen the observed effects of maximum duration of entitlement on job finding rates.

Finally, the type of unemployment benefit may have an impact on the effect on the job finding rate. As mentioned above, some countries employ two systems to provide benefits to unemployed individuals: an unemployment insurance system for individuals who typically have a strong labour market attachment (UI benefits) and a social welfare system for individuals who often have other problems in addition to unemployment (SA or UA benefits). The effect size in social welfare systems offering unemployment benefits with a known expiration date is, due to the participants' lower labour market attachment, expected to be less than the effect size in unemployment insurance systems with a known expiration date.

Why it is Important to do the Review

In order to reduce high unemployment levels, policy-makers may wish to reduce the generosity of the unemployment system either in amount (the replacement rate) or in maximum potential duration.

The positive correlation between unemployment benefit generosity in terms of the replacement rate and unemployment duration is well established at the empirical level (Layard et al., 2005). However, it may be politically intractable to lower the replacement rate, and there are indeed strong efficiency and equity arguments for having a reasonable value of unemployment benefits (Acemoglu & Shimer, 1999; Marimon & Zilibotti, 1999).

Search theory suggests that an increase in unemployment benefit generosity, in terms of maximum duration of benefit entitlement, has a negative impact on the job search activities of the unemployed increasing their unemployment duration.

There is clear evidence that the prospect of exhaustion of benefits results in a significantly increased incentive for finding work, although the effect is small (Filges

et al., 2013). Hence, shortening the benefit eligibility period may reduce the share of long and unproductive job searches somewhat. The conclusion in Filges et al. (2013) however leaves unanswered the question of by how much reducing the maximum unemployment benefit entitlement will decrease unemployment duration.

There are many empirical papers on the effect of maximum benefit entitlement on unemployed individuals (Caliendo, Tatsiramos and Uhlenhoff 2009; Bennmarker, Carling & Holmlund, 2007; Ham & Rea, 1987; Hunt, 1995; Katz & Meyer, 1990 and Lalive & Zweimüller, 2004), but the empirical research has not been summarized in a systematic review to obtain a clearer picture of the available evidence on the employment effect of reducing maximum duration of benefit entitlement. One paper provides a review of the literature on how incentives in unemployment insurance can be improved (Fredriksson and Holmlund 2006). However, it is not a systematic review and, furthermore, the authors do not make the important distinction between exits to employment and exits to other destinations such as such as other kinds of benefits or out of the labour force. Distinguishing between destinations is vital. As shown in Card, Chetty and Weber (2007), the exit rate from registered unemployment increases over 10 times more than the rate of re-employment at the expiration of benefits. The difference between the two measures arises because many individuals leave the unemployment register immediately after their benefits expire without returning to work.

There is a great deal of political interest in optimizing the unemployment benefit system so it balances the protection and distortion dimensions. The political interest is to reduce the unemployment level, to prevent exploitation of the unemployment benefit system and at the same time protect the unemployed individuals from the consequences of involuntary unemployment. It is therefore of great importance to establish the effect of reducing maximum duration of unemployment benefit entitlement on employment probabilities.

OBJECTIVES

The purpose of this review is to systematically uncover relevant studies in the literature that measure the effects of shortening the maximum duration of unemployment benefit entitlement on job finding rates, and to synthesize the effects in a transparent manner. As a secondary objective we will, where possible, investigate the extent to which the effects differ among different groups of unemployed such as high/low educated or men/women, and further explore from which point in the unemployment spell do unemployed individuals react to the length of benefit entitlement.

METHODOLOGY

Title registration

The title for this systematic review was approved in The Campbell Collaboration on 9. October 2012.

Criteria for including and excluding studies

Types of study designs

The proposed project will follow standard procedures for conducting systematic reviews using meta-analysis techniques. The study designs eligible for inclusion are:

- Controlled trials:
 - RCT - randomized controlled trial
 - QRCT - quasi-randomized controlled trial (i.e., participants are allocated by means such as alternate allocation, person's birth date, the date of the week or month, or alphabetical order)
 - NRCT - non-randomized controlled trial (i.e. 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 means such as time differences, location differences, decision-makers, or policy rules.

Study designs that use a well-defined control group are eligible. The main control or comparison condition is no change in maximum duration of benefit entitlement.

Non-randomized studies, where the reduction in maximum duration of benefit entitlement has occurred in the course of usual decisions outside the researcher's control must demonstrate pre-treatment group equivalence via matching, statistical controls, or evidence of equivalence on key risk variables (e.g., labour market conditions) and participant characteristics. These factors are outlined in section '***Assessment of risk of bias in included studies***' under the subheading of *Confounding*, and the methodological appropriateness of the included studies will be assessed according to the risk of bias model outlined in section '***Assessment of risk of bias in included studies***'.

Studies of the effect of reducing unemployment benefit entitlement typically are estimated on data collected from administrative registers or by questionnaires. Studies that use different data sources for treatment and control groups will not be eligible.

Only studies that use individual micro-data are eligible. Studies that rely on regional or national time series data are not eligible, even though micro-econometric estimates of individual effects merely provide partial information about the full impact of shortening the maximum duration of benefit entitlement (Calmfors, 1994; Calmfors, 1995).

We will include studies irrespective of their publication status, and their electronic availability.

Types of participants

We will include unemployed individuals who receive some sort of time limited benefit during their unemployment spell. 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). We will include participants receiving all types of unemployment benefits with a known exhaustion date. The only restriction is that the benefits must be related to being unemployed. We will therefore exclude individuals who only receive other types of benefits not related to being unemployed. We will include all unemployed participants regardless of age, gender, etc. who receive some sort of time limited benefit during their unemployment spell.

Types of interventions

The intervention is reduction in the maximum duration of entitlement of any kind of unemployment benefits. The benefits may be unemployment insurance (UI) benefits or they may be unemployment assistance (UA)/social assistance (SA). The only requirement is that the benefit must have a known expiration date. The UI benefit usually has a known time-limit whereas UA and SA usually are indefinite.

Unemployment benefits with an indefinite time limit or non-financial benefits will be excluded from this review.

Types of outcome measures

The objective is to determine whether reducing the maximum entitlement to unemployment benefits motivates unemployed individuals to find a job more quickly. Distinguishing between destinations is therefore vital. The primary outcome is exits to employment. Studies only looking at exits to other destinations such as other types of social benefits or non-employment and studies who do not distinguish between destinations are not eligible.

We will consider secondary outcomes in terms of the impact that reducing the maximum duration of entitlement of benefit has on the duration of re-employment and on income. This will be done in order to obtain a clearer picture of the effect that reducing the maximum entitlement of unemployment benefit has on the quality of the job. If the duration of re-employment or the wage is low, this could indicate that reducing entitlement forces unemployed individuals to find jobs that do not match their qualifications and therefore they may return to unemployment quickly.

Primary outcomes refer to employment status:

- a) exit rate, measured as a hazard 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

Duration of follow-up

Outcomes measured as hazard ratios may be reported as an overall effect on the hazard ratio, or may be reported separate for various unemployment duration intervals. All time points reported will be considered.

Types of settings

All types of settings are eligible.

Search strategy

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 are applied in the searches⁶.

The following bibliographic databases will be searched:

- Business Source Elite
- EconLit
- PsycINFO
- SocIndex
- Science Citation Index
- Social Science Citation Index
- The Cochrane Library (Cochrane reviews, other reviews, clinical trials)
- International Bibliography of the Social Sciences
- IDEAS/Economist Online/Social Care Online ⁷
- Dissertations and Theses (Aka Dissertation Abstracts)
- Theses Canada

An example of the search strategy for Business Source Elite is listed in Appendix 1.4. The strategy will be modified for the different databases. We will report full details of the modifications in the completed review.

Searching other resources

Grey literature

Additional searches will be made by means of Google (including Google Scholar) and we will check the first 150 hits. OpenSIGLE will be used to search for European grey

⁶ The first part of the search period (until March 2011) will be covered by re-examining results of the searches for an earlier review (Filges et al., 2013) that used an identical search strategy.

⁷ The search strategy will be modified for these databases as the search interfaces do not allow complex searching. Even though these databases contain similar references, we will search both with the aim of performing as thorough a search as possible.

literature (<http://opensigle.inist.fr/>). Copies of relevant documents will be made and we will record the exact URL and date of access for each relevant document.

Websites of the following private independent research institutes and economic networks will be searched:

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)

MDRC – the Manpower Demonstration Research Corporation – (www.mdrc.org)

CESifo – the cooperation between CES (Center for Economic Studies) and IFO (Institute for Economic Research) – (www.cesifo-group.de/portal/page/portal/ifoHome) are all covered via IDEAS.

In addition we will look into the following sites:

Danish Economic Councils (www.dors.dk)

OECD - the Organisation for Economic Co-operation and Development (www.oecd.org)

IMF - The International Monetary Fund (www.imf.org)

AIECE - Association of European Conjuncture Institutes (www.aiece.org)

ESRC - Economic Social Research Council (www.esrc.ac.uk)

Copenhagen Economics (www.copenhageneconomics.com)

SSRN – Social Science Research Network (www.ssrn.com) will also be searched to uncover potential preprint discussion papers.

Unpublished theses and dissertations will be searched through the databases:

Theses and dissertations and Theses Canada.

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

Hand searching

Reference lists of included studies and reference lists of relevant reviews will be searched. “The Journal of Labor Economics” and “Labour Economics” will be hand searched for the year 2014 and the available issues of 2015.

Snowballing

Reference lists of included studies and relevant reviews will be searched for potential new literature.

Personal contacts

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

Description of methods used in primary research

We expect that a proportion of the studies we locate will have been conducted without randomisation of participants, since there is no firm tradition for RCTs in labour market research. This stems among other things from some degree of scepticism towards randomisation of participants due to ethical concerns about random allocation of services. The central problem in these studies without randomisation of participants is the identification of the causal effect. Many studies use variation in benefit rules or legislative changes of the maximum entitlement period.

A frequently adopted policy is to extend the maximum benefit period when labour market conditions are expected to deteriorate. Several studies use this policy to estimate the effect on unemployment duration (Ham & Rea, 1987; Hunt, 1995 and Katz & Meyer, 1990). It is however problematic to rely on such a rule as these changes are, like any other policy rule, purposeful action. If the determinants of the change are not accounted for it will yield biased estimates. Part of the effect will be due to the changed labour market conditions that lead to the change in entitlement in the first place. A more recent study (Lalive & Zweimüller, 2004) uses extended benefit entitlement in Austria and adopts four different identification strategies in order to disentangle the causal effect of extended benefit entitlement from the impact of changed labour conditions. The extended benefit entitlement was enacted to mitigate the labour market problems in certain regions and for certain subgroups of workers. The extension was therefore limited to job seekers aged 50 or more, living in certain regions, for a limited time period (the rules were reformed after few years), implying that there may be many non-entitled workers who are quite similar to entitled individuals. Their different identification strategies account for time trends using difference-in-differences-in-difference strategies and they choose

different subgroups of treated where no idiosyncratic shocks are expected. They consider the policy of extended benefit entitlement as ‘exogenous’⁸ (and thereby usable to estimate the causal effect) when it can be reasonably argued that the treated individuals are not subject to idiosyncratic shocks during the observation period (see Lalive & Zweimüller, 2004 for further details).

The same concerns of exogeneity apply to all legislative changes to the maximum entitlement period. For example the study by van Ours and Vodopivec (2006), exploit reforms of the Slovenian unemployment insurance system in 1998. The reform reduced the maximum duration of benefits, roughly by half for most groups of recipients. To identify the effect, they adopt a difference-in-difference strategy and compare the probability of entering employment before and after the reform for those affected with the job finding probability for those who were not affected. To avoid bias due to expectations of the reform affecting inflows from employment to unemployment, they only consider data for the period of 2 months before and 2 months after the reform’s introduction. The authors further argue that it is a credible identification strategy, as the reforms introduced variation in potential benefit duration unrelated to the state of the labour market and furthermore, changed potential benefit duration differently for different groups of unemployed.

In Caliendo, Tatsiramos & Uhlendorff (2009) the analysis is based on a regression discontinuity design. The identification strategy relies on a sharp discontinuity in the maximum duration of unemployment benefits at the age of 45 in Germany. Comparing unemployed who are just below the age threshold with unemployed just above the corresponding age gives a measure of the effect of maximum duration of benefits. Likewise is the analysis in Lalive (2008) based on a regression discontinuity design, using discontinuities in the potential benefit duration at age 50 and across regions in Austria.

Criteria for determination of independent findings

We will take into account the unit of analysis of the studies to determine to 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.

⁸ The study authors use quotations marks.

Cluster randomised trials

Cluster randomised trials included in this review will be checked for consistency in the unit of allocation and the unit of analysis, as statistical analysis errors can occur when they are different. When appropriate analytic methods have been used, we will meta-analyse effect estimates and their standard errors (Higgins & Green, 2011). In cases where study investigators have not applied appropriate analysis methods that control for clustering effects, we will estimate the intra-cluster correlation (Donner, Piaggio, & Villar, 2001) and correct standard errors.

Multiple interventions groups and multiple interventions per individuals

Studies with multiple intervention groups with different individuals will be included in this review. To avoid problems with dependence between effect sizes we will apply robust standard errors (Hedges, Tipton, & Johnson, 2010). However, simulation studies show that this method needs around 20-40 studies included in the data synthesis (Hedges et al., 2010). If this number cannot be reached we will use a synthetic effect size (the average) in order to avoid dependence between effect sizes. This method provides an unbiased estimate of the mean effect size parameter but overestimates the standard error. Random effects models applied 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 (e.g., the 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. The choice of which estimate to include will be based on our risk of bias assessment. We will choose the estimate that we judge to have the least risk of bias (primarily, selection bias and in case of equal scoring the incomplete data item will be used).

Multiple studies using the same sample of data

In some cases, several studies may have used the same sample of data or some studies may have used only a subset of a sample used in another study. We will review all such studies, but in the meta-analysis we will only include one estimate of the effect from each sample of data. This will be done to avoid dependencies between the “observations” (i.e. the estimates of the effect) in the meta-analysis. The choice of which estimate to include will be based on our risk of bias assessment of the studies. We will choose the estimate from the study that we judge to have the

least risk of bias (primarily, selection bias). If two (or more) studies are judged to have the same risk of bias and one of the studies (or more) uses a subset of a sample used in another study or studies, we will include the study using the full set of participants.

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 measurements at a similar time point. As a general guideline, these will be grouped together according to length of unemployment duration as follows: 1) 0 to less than 6 months, 2) 6 months to 12 months, 3) more than 1 year. However, should the studies provide viable reasons for an adjusted choice of relevant and meaningful duration intervals for the analysis of outcomes, we will adjust the grouping.

Details of study coding categories

Selection of studies and data extraction

Under the supervision of review authors, two review team assistants will first independently screen titles and abstracts to exclude studies that are clearly irrelevant. Studies considered eligible by at least one assistant or studies where there is insufficient information in the title and abstract to judge eligibility, will be retrieved in full text. The full texts will then be screened independently by two review team assistants under the supervision of the review authors. Any disagreement of eligibility will be resolved by the review authors. Exclusion reasons for studies that otherwise might be expected to be eligible will be documented and presented in an appendix.

The study inclusion criteria will be piloted by the review authors (see Appendix 1.1). The overall search and screening process will be illustrated in a flow-diagram. None of the review authors will be blind to the authors, institutions, or the journals responsible for the publication of the articles.

Two review authors will independently code and extract data from included studies. A coding sheet will be piloted on several studies and revised as necessary (see Appendix 1.2 and 1.3). Disagreements will be resolved by consulting a third review author with extensive content and methods expertise. Disagreements resolved by a third reviewer will be reported. 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,

outcomes, and results. Extracted data will be stored electronically. Analysis will be conducted using RevMan5, SAS and Stata software.

Assessment of risk of bias in included studies

We will assess the risk of bias using a model developed by Prof. Barnaby Reeves in association with the Cochrane Non-Randomised Studies Methods Group (Reeves, Deeks, Higgins, & Wells, 2011).⁹ This model is an extension of the Cochrane Collaboration's risk of bias tool and covers risk of bias in non-randomised studies that have a well-defined control group.

The extended model is organised and follows the same steps as the risk of bias model according to the 2008-version of the Cochrane Hand book, chapter 8 (Higgins & Green, 2008). The extension to the model is explained in the three following points:

1) The extended model 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 confounding¹⁰. 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 1.3). This assessment informs the final risk of bias score for confounding.

2) Another feature of non-randomised studies that make them at high risk of bias is that they need not have a protocol in advance of starting the recruitment process. 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, e.g., choice of method of model fitting, potential confounders considered / included. In addition, the model includes two separate yes/no items asking reviewers whether they think the researchers had a pre-specified protocol and analysis plan.

⁹ 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).

¹⁰ See next page for an explanation of the terms selection bias and confounding.

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 data synthesis.

Risk of bias judgement items

The risk of bias model used in this review is based on 9 items (see Appendix 1.3).

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)

In the 5-point scale, 1 corresponds to Low risk of bias and 5 corresponds to High risk of bias. A score of 5 on any of the items assessed on the 5-point scale translates to a risk of bias so high that the findings will not be considered in the data synthesis (because they are more likely to mislead than inform).

Confounding

An important part of the risk of bias assessment of non-randomised studies is consideration of how the studies deal with confounding factors (see Appendix 1.3). Selection bias is understood as systematic baseline differences between groups which 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 represent different approaches to dealing with selection problems under different assumptions, and consequently require different types of data. There can be particularly great variations in how different designs deal with selection on unobservables. The “adequate” method depends on the model generating participation, i.e. assumptions about the nature of the process by which participants are selected into a programme. A major difficulty in estimating causal effects of the maximum duration of benefit entitlement is the potential endogeneity of the change to benefit rules stemming from the policy process that leads to the change.

The determinants of the change are often labour market conditions and if not accounted for it will yield biased estimates.

As there is no universal correct way to construct counterfactuals for non-randomised designs, we will look for evidence that identification is achieved, and that the authors of the primary studies justify their choice of method in a convincing manner by discussing the assumption(s) leading to identification (the assumption(s) that make it possible to identify the counterfactual). Preferably the authors should make an effort to justify their choice of method and convince the reader that the only difference between an individual with a short maximum benefit period and an individual with a longer maximum benefit period is exactly the difference in length of maximum benefit period and that the source of difference between their entitlement status is not endogenous to the individuals’ exit rate to employment. The judgement is reflected in the assessment of the confounder unobservables in the list of confounders considered important at the outset (see Appendix 1.3).

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

Importance of pre-specified confounding factors

The motivation for focusing on age, gender, education, ethnicity, labour market conditions and unemployment duration is given below.

The motivation for focusing on age, gender, education and ethnicity is that they are the major determinants of the risk of being unemployed (Layard et al., 2005).

Concerning unemployment duration, most studies find that the genuine duration dependence is negative, i.e. the longer the unemployment spell the smaller is the chance of finding a job¹¹ (see Serneels, 2002, for an overview). If the study does not disentangle the effect of shortening the maximum benefit period from the negative duration dependence the effect will be biased.

Another potential source of bias is differences in labour market conditions. If the study, for example, explores changes in the maximum duration of benefit entitlement over time or space as the source of variation, 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.

Assessment

At least two review authors will independently assess the risk of bias for each included study. Any disagreements will be resolved by a third reviewer with content and statistical expertise and will be reported. We will report the risk of bias assessment in risk of bias tables for each included study in the completed review.

Measures of treatment effect

We expect that the primary treatment effect will be measured either as the relative exit rate from unemployment to employment (measured as hazard ratio) or as the difference in probability of employment (measured as risk difference). Alternatively it may be measured directly as difference in mean duration (time to employment). For such a continuous outcome, 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.

Our main interest is to include studies in a meta-analysis where hazard ratios and variances are either reported or are calculable from the available data. Should sufficient effect sizes be available, we will perform a meta-analysis on the individual

¹¹ 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. This "genuine" duration dependence should not be confused with sorting which is another mechanism.

included studies using the log hazard ratio and variance. We will report the 95% confidence intervals.

The hazard ratio measures the proportional change in hazard rates between unemployed individuals who have a short maximum benefit period and unemployed individuals who have a longer maximum benefit period. The hazard rate is defined as the event rate (in the present context, the event is finding a job) at time t conditional on survival (staying unemployed) until time t or later. A hazard rate is constructed as follows:¹²

The length of an unemployment spell for an unemployed individual (in the present context the length of stay in the unemployment system until finding a job) is a realization of a continuous random variable T . In continuous time, the hazard rate $\theta(t)$ is defined as:

$$\theta(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)},$$

where the cumulative distribution function of T is:

$$F(t) = \Pr(T < t)$$

and the probability density function is:

$$f(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t} = \frac{dF(t)}{dt}.$$

$F(t)$ is also known in the survival analysis literature as the failure function and in the present context failure means finding a job. $S(t)$ is the survivor function:

$$S(t) \equiv \Pr(T \geq t) = 1 - F(t);$$

t is the elapsed time since entry to the state (since the individual entered the unemployment system).

Introducing covariates the hazard rate becomes:

$$\theta(t|x(t,s)) = \lim_{\Delta t \downarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t, x(t,s))}{\Delta t},$$

where $x(t,s)$ is a vector of personal characteristics that may vary with unemployment duration (t) or with calendar time (s).

A proportional hazard rate is given by:

$$\theta(t|x) = \theta_0(t) * \exp(x'\beta),$$

¹² The following description of hazard rates is based on Jenkins (2005) and van den Berg (2001).

where $\theta_0(t)$ is the baseline hazard, $\exp(x'\beta)$ is a scale function of the vector x of personal characteristics (and a treatment indicator) and β is a vector of estimated parameters.

The baseline hazard is typically not completely specified; often the hazard function is modelled as piecewise constant. Thus whether the shape of the hazard generally increases or decreases with survival time is left to be estimated from the data, rather than specified a priori.

In the description of the hazard rate it is, so far, implicitly assumed that all relevant differences between individuals can be summarized by observed explanatory variables. But if there are unobservable differences, e.g. motivation and 'ability' (in the literature termed unobserved heterogeneity) and these differences are ignored, the estimated parameters will be biased towards zero. It is therefore common to control for both observed factors given by the vector x as well as unobserved factors, i.e. unobserved heterogeneity. The hazard rate, including unobserved heterogeneity, is now given by:

$$\theta(t|x, v) = \theta_0(t) * \exp(x'\beta)v,$$

where v represents factors unobserved to the researcher and independent of x . It is necessary to assume the distribution of v has a shape where the right-hand tail of the distribution is not too fat and whose functional form is summarized in terms of only a few key parameters, in order to estimate those parameters with the data available. The unobserved components are typically assumed to follow a discrete distribution with two (or more) points of support.

The acceptable outcome measurement frequency for calculating hazard ratios in this review will be three months or less. A study reporting only outcomes measured on time intervals of more than three months will not be included in the meta-analysis using hazard ratios.

Studies providing estimates of hazard ratios and variances typically base the estimation on the maximum likelihood method¹³. The principle of maximum likelihood is relatively straightforward. The likelihood function, regarded as a function of the parameters of the model, is the joint density of the observations. The maximum likelihood estimator yields a choice of the estimator as the value for the parameter that makes the observed data most probable.

¹³ The following description of estimation is based on Lancaster, 1990.

Ignoring unobserved heterogeneity, the contribution to the likelihood for complete observations is given by the conditional density function of t :

$$f(t|x) = \theta(t|x) \exp\left(- \int \theta(s|x) ds\right)$$

and for censored observations:

$$S(t|x) = \exp\left(- \int \theta(s|x) ds\right)$$

The likelihood function is:

$$L = f(t|x)^d S(t|x)^{1-d}$$

where $d=1$ for complete observations and $d=0$ for censored observations. Often it is convenient to maximise the logarithm of the likelihood function rather than the likelihood function and the same results are obtained since $\log L$ and L attain the maximum at the same point.

The log likelihood function to maximize with respect to the parameters of the model is:

$$\log L = d \log f(t|x) + (1-d) \log S(t|x) = d \log \theta(t|x) - \int \theta(s|x) ds$$

Introducing unobserved heterogeneity with the random components assumed to follow a discrete distribution with two points of support ($v_1, v_2, \Pr(v_1) = \pi_1, \Pr(v_2) = \pi_2$) the log likelihood function becomes:

$$\log L = \left(d \log \theta(t|x) - \int \theta(s|x) ds \right) \pi_1 + \left(d \log \theta(t|x) - \int \theta(s|x) ds \right) \pi_2$$

If hazard ratios and variances are not reported, log hazard ratios and variances will be computed directly using the observed number of events and log rank expected number of events if available (Parmar, Torri, & Stewart, 1998).

The log hazard ratio will be calculated as: $\log(HR) = \log((Oa/Ea)/(Ob/Eb))$, where Oa and Ob is the number of observed events in each group and Ea and Eb is the number of expected events assuming a null hypothesis of no difference in survival.

The standard error of the log hazard ratio will be calculated as $\sqrt{(1/Ea + 1/Eb)}$.

Alternatively log hazard ratios and variances will be computed 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).

For continuous outcomes (such as mean duration), effects sizes with 95 % confidence intervals will be calculated, where means and standard deviations are available. If means and standard deviations are not available, we will calculate

standardized mean differences (SMD) from F-ratios, t-values, chi-squared values and correlation coefficients, where available, using the methods suggested by Lipsey & Wilson (2001). Hedges' *g* will be used for estimating SMDs.

The review authors will not request information from the principal investigators if not enough information is provided to calculate an effect size and standard error due to the time span of studies (the time span between the earliest we know of and the latest is 30 years). If missing summary data (e.g. valid Ns, means and standard deviations) cannot be derived, the study results will be reported in as much detail as possible.

For secondary outcomes, duration of re-employment may be measured as hazard rates in which case the effect size will be measured as log hazard ratios or relative risk ratio. We will report the 95% confidence intervals. Alternatively it may be measured directly as mean duration. Income may be measured as the mean income at different time points or during different time periods. Hedges' *g* will be used for estimating SMDs where scales measure the same outcomes in different ways. We will report the 95% confidence intervals.

RevMan 5.0, Excel and Stata 10.0 software will be used for the statistical analyses.

Dealing with missing data and incomplete data

The reviewers will assess missing data rates in the included studies in accordance with the risk of bias tool used (see section *Assessment of risk of bias in included studies*).

The reviewers will record information on intention to treat analysis (ITT). We will perform sensitivity analysis to examine influences on effects in studies using ITT analysis vs. studies not using ITT analysis.

Statistical procedures and conventions

We will follow standard procedures for conducting systematic reviews using meta-analysis techniques. The overall data synthesis will be conducted where effect sizes are available or can be calculated, and where studies are similar in terms of the outcome measured. Meta-analysis of both primary and secondary outcomes will be conducted on each metric (as outlined in section '*Types of outcomes measures*') separately.

As different computational methods may produce effect sizes that are not comparable, we will be transparent about all methods used in the primary studies (research design and statistical analysis strategies) and use caution when synthesizing effect sizes. Special caution will be taken concerning studies using instrumental variables (IV) and regression discontinuity (RD) to estimate a local average treatment effect (LATE) (Angrist & Pischke, 2009). These will be included, but may be subject to a separate analysis depending on the comparability between the LATE's and the effects from other studies. We will in any case check the sensitivity of our results to the inclusion of IV and RD studies. In addition we will discuss the limitation in generalisation of results obtained from these types of studies.

When the effect sizes used in the data synthesis are hazard ratios, they will be log transformed before being analysed. The reason is that ratio summary statistics all have the common feature that the lowest value that they can take is 0, that the value 1 corresponds with no intervention effect, and the highest value that a hazard ratio can ever take is infinity. This number scale is not symmetric. The log transformation makes the scale symmetric: the log of 0 is minus infinity, the log of 1 is zero, and the log of infinity is infinity.

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.

As the intervention deal with diverse populations of participants (from different countries, facing different labour market conditions etc.), and we therefore expect heterogeneity among primary study outcomes, all analyses of the overall effect will be inverse variance weighted using random effects statistical models that incorporate both the sampling variance and between study variance components into the study level weights. Random effects weighted mean effect sizes will be calculated using 95% confidence intervals and we will provide a graphical display (forest plot) of effect sizes. Graphical displays for meta-analysis performed on ratio scales sometimes use a log scale, as the confidence intervals then appear symmetric. This is however not the case for the software Revman 5 which we plan to use in this review. The graphical displays using hazard ratios and the mean effect size will be reported as a hazard ratio. Heterogeneity among primary outcome studies will be assessed with Chi-squared (Q) test, and the I-squared, and τ -squared statistics (Higgins, Thompson, Deeks, & Altman, 2003). Any interpretation of the Chi-squared test will be made cautiously on account of its low statistical power.

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).

We expect that several studies have used the same sample of data. We will review all such studies, but in the meta-analysis we will only include one estimate of the effect from each sample of data. This will be done to avoid dependencies between the “observations” (i.e. the estimates of the effect) in the meta-analysis. 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, with particular attention paid to selection bias.

We anticipate that several studies provide results separated by for example age and/or gender. We will include results for all age and gender groups. To take into account the dependence between such multiple effect sizes from the same study, we will apply robust standard errors (Hedges et al., 2010). An important feature of this analysis is that the results are valid regardless of the weights used. For efficiency purposes, we will calculate the weights using a method proposed by Hedges et al (2010). This method assumes a simple random-effects model in which study average effect sizes vary across studies (τ^2) and the effect sizes within each study are equicorrelated (ρ). The method is approximately efficient, since it uses approximate inverse-variance weights: they are approximate given that ρ is, in fact, unknown and the correlation structure may be more complex. We will calculate weights using estimates of τ^2 , setting $\rho = 0.80$ and conduct sensitivity tests using a variety of ρ values; to assess if the general results and estimates of the heterogeneity is robust to the choice of ρ .

This robust standard error method uses degrees of freedom based on the number of studies (rather than the total number of effect sizes). Simulation studies show that this method needs around 20-40 studies included in the data synthesis (Hedges et al., 2010). If this number cannot be reached we will conduct a data synthesis where we use a synthetic effect size (the average) in order to avoid dependence between effect sizes.

Moderator analysis and investigation of heterogeneity

We will investigate the following factors with the aim of explaining potential observed heterogeneity: 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), labour market conditions (good/bad), type of unemployment benefit (UI

or SA/UA), whether alternative benefits are available and if compulsory activation is part of the system.

If the number of included studies is sufficient and given there is 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.

If there are a sufficient number of studies we will apply robust standard errors and calculate the weights using a method proposed by Hedges et al. (2010). This technique calculates standard errors using an empirical estimate of the variance: it does not require any assumptions regarding the distribution of the effect size estimates. The assumptions that are required to meet the regularity conditions are minimal and generally met in practice. Simulation studies show that both confidence intervals and p-values generated this way typically reflect the correct size in samples, requiring between 20-40 studies. This more robust technique is beneficial because it takes into account the possible correlation between effect sizes separated by the covariates within the same study and allows all of the effect size estimates to be included in meta-regression. We will calculate weights using estimates of τ^2 , setting $\rho = 0.80$ and conduct sensitivity tests using a variety of ρ values; to assess if the general results and estimates of the heterogeneity is robust to the choice of ρ .

We will report 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 cautiously drawn and will not solely be based on significance tests. The magnitude of the coefficients and width of the confidence intervals will be taken into account as well. Otherwise, single factor subgroup analysis will be performed. The assessment of any difference between subgroups will be based on 95% confidence intervals. 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; see Oxman & Guyatt (1992). We will therefore use within study differences where possible.

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

Sensitivity analysis

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

Sensitivity analysis will be carried out to ensure the consistency of results from studies using a reduction and studies using an extension of maximum benefit entitlement to estimate the effect.

Further sensitivity analyses with regard to research design and statistical analysis strategies in the primary studies will be an important element of the analysis to ensure that different methods produce consistent results.

Assessment of reporting bias

Reporting bias refers to both publication bias and selective reporting of outcome data and results. Here, we state how we will assess publication bias.

We will use funnel plots for information about possible publication bias if we find sufficient 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.

Treatment of qualitative research

We do not plan to include qualitative research.

REFERENCES

- Acemoglu, D. and Shimer, R. (1999). Efficient unemployment insurance. *Journal of Political Economy*, 107, page 893-928.
- Angrist, J.D., & Pischke, J.S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

- Benmarker, H., Carling, K. and Holmlund, B. (2007). Do Benefit Hikes Damage Job Finding? Evidence from Swedish Unemployment Insurance Reforms. *Labour*, 21(1), 85–120
- 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., Chetty, R. and Weber, A. (2007). The Spike at Benefit Exhaustion: Leaving the Unemployment System or Starting a New Job. *IZA DP No. 2590*.
- Card, D. and Riddell, W.C. (1993). A Comparative Analysis of Unemployment in Canada and the United States. In: Card, D. and Freeman, R.B. (eds.). *Small Differences That Matter: Labor Markets and Income Maintenance in Canada and the United States*, 149-190. University of Chicago Press. Available at: http://economics.ubc.ca/files/2013/05/pdf_paper_craig-riddell-comparative-analysis-unemployment-canada.pdf
- Caliendo, M., Tatsiramos, K. and Uhlendoff, A. (2009). Benefit Duration, Unemployment Duration and Job Match Quality: A regression-Discontinuity Approach. *IZA DP No. 4670*.
- Donner, A., Piaggio, G. and Villar, J. (2001). Statistical methods for the meta-analysis of cluster randomized trials. *Statistical Methods in Medical Research*, No. 10, page 325-38.
- Feldstein, M (2005). Rethinking Social Insurance. *The American Economic Review*, Vol. 95, No. 1, pp. 1-24
- Filges, T. and Hansen, A.T. (2015). The threat of active labor market programs: A systematic review. Manuscript submitted.
- Filges, T. & Larsen, B. (2000). Does active labour market policy reduce unemployment when search is endogenous? *WP 14-2000, Department of Economics, Copenhagen Business School*.
- Filges, T., Geerdsen, L.P., Knudsen, AS.D., Jørgensen, AM.K. and Kowalski, K. (2013). Unemployment Benefit Exhaustion: Incentive Effects on Job Finding Rates: A Systematic Review. *Campbell Collaboration*, 2013:4.
- Filges, T., Smedslund, G., Knudsen, AS.D. and Jørgensen, AM.K. (2015). Active Labour Market Programme Participation for Unemployment Insurance Recipients: A Systematic Review. *Campbell Collaboration*, 2015:2.

- Fredriksson, P. and Holmlund, B. (2006). Improving incentives in unemployment insurance: A review of recent research. *Journal of economic surveys*, vol. 20, No. 3.
- Gustafson, C, and P. Levine (1997): Less-skilled workers, welfare reform and the Unemployment Insurance, Working Paper, University of California at Berkeley, Dept. of Economic.
- Ham, J., Rea, S., 1987. Unemployment insurance and male unemployment duration in Canada. *Journal of Labor Economics* 5, 325– 353.
- Hedges, L. V. (2007). Meta-analysis. In: Rao, C.R. (ed.). *The Handbook of Statistics*, 919-53. Amsterdam: Elsevier.
- Hedges, L.V., Tipton, E. & Johnson, M.C. (2010). Robust variance estimation in meta regression with dependent effect size estimates. *Research Synthesis Methods* 2010 (1), 39-65.
- Hedges, L.W. and Pigott, T.D., (2004). The Power of Statistical Tests for Moderators in Meta-Analysis. *Psychological Methods*, Vol. 9, No. 4, page 426–445.
- Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), page 557-60.
- Higgins, J.P.T., & Green, S. (eds.) (2008). *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley-Blackwell.
- 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.
- Hunt, J., 1995. The effect of unemployment compensation on unemployment duration in Germany. *Journal of Labor Economics* 13, 88–120.
- International Labour Office (1990), *Employment, unemployment and underemployment. An ILO Manual on Concepts and Methods*, Geneva, International Labour Office.
- Jenkins, S. P. (2005), Survival Analysis. Lecture Notes, 18 July. <http://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf>
- Katz, L., Meyer, B., 1990. The impact of the potential duration of unemployment benefits on the duration of unemployment. *Journal of Public Economics* 41, 45– 72.
- Kluve, J., Lehmann, H. and Schmidt, C.M. (1999). Active Labor Market Policies in Poland: Human Capital Enhancement, Stigmatization or Benefit Churning. *Journal of Comparative Economics*, 27, page 61-68.

- Kluve, J., Card, D., Fertig, M., Góra, M., Jacobi, L., Jensen, P., Leetmaa, R., Nima, L., Patacchini, E., Schaffner, S., Schmidt, C.M., van der Klaauw, B. and Weber, A. (2007). *Active Labor Market Policies in Europe: Performance and Perspectives*. Springer.
- Konstantopoulos, S. (2006). Fixed and Mixed Effects Models in Meta-Analysis. *IZA DP no. 2198*.
- Lalive, R. (2008). How do extended benefits affect unemployment duration? A regression discontinuity approach. *Journal of Econometrics*, 142, 785–806.
- Lalive, R. and Zweimüller, J. (2004) Benefit entitlement and unemployment duration - The role of policy endogeneity. *Journal of Public Economics* 88 , 2587–2616
- Lancaster, T. (1990). The econometric analysis of transition data. *Econometric Society Monographs No. 17*, Cambridge: Cambridge University Press.
- Layard, R., Nickell, S. and Jackman, R. (2005). *Unemployment*. Oxford University Press.
- Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. *Applied Social Research Methods Series*, v. 49.
- Marimon, R. and Zilibotti, F. (1999). Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits. *Economic Journal*, 109, page 266–291.
- Mortensen, D. (1977). Unemployment Insurance and Job Search Decisions. *Industrial and Labor Relations Review*, 30, page 505–517.
- Mortensen, D. T (1987). Job Search and Labor Market Analysis. *Handbook of Labor Economics Vol. 2*, North-Holland
- 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).
- Parmar, M.K.B., Torri, V. and Stewart L. (1998). Extracting summary statistics to perform meta analyses of the published literature for survival endpoints. *Statistics in Medicine*, 17, page 2815-2834.
- Pissarides, Christopher A. (2000). *Equilibrium Unemployment Theory*, 2nd ed., Cambridge (MA): MIT Press.
- Reeves, B. C , Deeks, J. J., Higgins, J. P. T., & Wells G. A. (2011). Including non-randomized studies. Unpublished manuscript.

Rothstein, J. (2011). Unemployment Insurance and Job Search in the Great Recession. *NBER Working Paper No. 17534*.

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.

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., and Song, F. (2000). *Methods of meta-analysis in medical research*. New York: Wiley.

Van den Berg, G.J. (1990). Nonstationarity in Job Search Theory. *The Review of Economic Studies, 57(2)*, pp. 255-277.

Van den Berg, G.J. (2001). Duration Models: Specification, Identification, and Multiple Durations. In: Heckman, J.J. and Leamer, E. (eds.), *Handbook of Econometrics, Volume V*. Amsterdam: North-Holland.

van Ours, J. C. & Vodopivec, M. (2006). How shortening the potential duration of unemployment benefits affects the duration of unemployment: Evidence from a natural experiment. *Journal of Labor Economics, 24*, 351-378.

REVIEW AUTHORS

Lead review author: The lead author is the person who develops and coordinates the review team, discusses and assigns roles for individual members of the review team, liaises with the editorial base and takes responsibility for the on-going updates of the review.

Name:	Trine Filges
Title:	Senior Researcher
Affiliation:	SFI-Campbell
Address:	Herluf Trollesgade 11
City, State, Province or County:	Copenhagen
Postal Code:	1052
Country:	Denmark
Phone:	45 33480926
Email:	tif@sfi.dk

Co-author(s):

Name:	Anders Bruun Jonassen
Title:	Researcher
Affiliation:	SFI-Campbell
Address:	Herluf Trollesgade 11
City, State, Province or County:	Copenhagen
Postal Code:	1052
Country:	Denmark
Phone:	45 69 77 33
Email:	abj@sfi.dk

Name:	Anne Marie Klint Jørgensen
Title:	Librarian/Information Specialist
Affiliation:	SFI-Campbell
Address:	Herluf Trollesgade 11
City, State, Province or County:	Copenhagen
Postal Code:	1052
Country:	Denmark
Phone:	45 33480868
Email:	amk@sfi.dk

ROLES AND RESPONSIBILITIES

- Content: Trine Filges and Anders Bruun Jonassen
- Systematic review methods: Trine Filges
- Statistical analysis: Trine Filges and Anders Bruun Jonassen
- Information retrieval: Anne Marie Klint Jørgensen

SOURCES OF SUPPORT

Internal funding: SFI-Campbell

External funding: None

DECLARATIONS OF INTEREST

None known

PRELIMINARY TIMEFRAME

We plan to submit a draft review: within 18 months of protocol approval

PLANS FOR UPDATING THE REVIEW

Trine Filges will be responsible for updating the review every second year when completed.

AUTHOR DECLARATION

Authors' responsibilities

By completing this form, you accept responsibility for preparing, maintaining and updating the review in accordance with Campbell Collaboration policy. The Campbell Collaboration will provide as much support as possible to assist with the preparation of the review.

A draft review must be submitted to the relevant Coordinating Group within two years of protocol publication. If drafts are not submitted before the agreed deadlines, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to de-register the title or transfer the title to alternative authors. The Coordinating Group also has the right to de-register or transfer the title if it does not meet the standards of the Coordinating Group and/or the Campbell Collaboration.

You accept responsibility for maintaining the review in light of new evidence, comments and criticisms, and other developments, and updating the review at least once every five years, or, if requested, transferring responsibility for maintaining the review to others as agreed with the Coordinating Group.

Publication in the Campbell Library

The support of the Coordinating Group in preparing your review is conditional upon your agreement to publish the protocol, finished review, and subsequent updates in the Campbell Library. The Campbell Collaboration places no restrictions on publication of the findings of a Campbell systematic review in a more abbreviated form as a journal article either before or after the publication of the monograph version in *Campbell Systematic Reviews*. Some journals, however, have restrictions that preclude publication of findings that have been, or will be, reported elsewhere and authors considering publication in such a journal should be aware of possible conflict with publication of the monograph version in *Campbell Systematic*

Reviews. Publication in a journal after publication or in press status in *Campbell Systematic Reviews* should acknowledge the Campbell version and include a citation to it. Note that systematic reviews published in *Campbell Systematic Reviews* and co-registered with the Cochrane Collaboration may have additional requirements or restrictions for co-publication. Review authors accept responsibility for meeting any co-publication requirements.

I understand the commitment required to undertake a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by: Trine Filges

Date:

1 Appendices

1.1 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 some kind of benefit during their unemployment?
Yes - include
No – if no then stop here and exclude
Uncertain - include

Question 1 guidance:

This includes all types of unemployment benefits both unemployment insurance benefits, unemployment assistance benefits and social assistance benefits.

2. Does the study focus on time limits in the unemployment benefit eligibility period or exhaustion of unemployment benefits or entitlement to

unemployment benefits or maximum duration of unemployment benefits
etc.?

Yes - include

No – if no then stop here and exclude

Uncertain - include

Question 2 guidance:

The intervention is a change in the maximum duration of any kind of unemployment benefit with a known expiration date. This intervention can be referred to in different ways.

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 5 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) or studies who do not distinguish between destinations will not be included.

1.2 DATA EXTRACTION

Language
Journal
Year
Country
Time period covered by data
Type of unemployment scheme (UI, social benefit other (specify))
Participation characteristics (age, gender, education, ethnicity, eligibility requirements for benefits)
Benefit level/replacement rate
Labour market conditions (unemployment rate and/or vacancy rate)
Benefit level/replacement rate available after exhaustion if any
Is compulsory activation part of the system? If yes, describe the elements of the programme (education, work, training, self-employment, job search assistance)
Maximum duration of unemployment benefits
Type of data used (register, questionnaire, other (specify))
Sampling frequency
Time interval the outcome measure is based on (if different from sampling frequency)
Is there correction for unobserved heterogeneity? If yes, how?
Sample size (Treatment/control)
Is there correction for censoring (yes/no)

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	Source	Blind (outcome assessors) ?	Pg# & notes
1		Info from: Other samples This sample Unclear Info provided:	Dichotomy Continuous Time-to-event	High score or event is Positive Negative Can't tell	Questionnaire Admin data Other (specify) Unclear	Yes No Can't tell	

* Repeat as needed

OUT COME DATA

DICHOTOMOUS OUTCOME DATA

OUTCOME	TIME POINT (s) (record exact time from beginning of unemployment, there may be more than one, record them all)	SOURCE	VALID Ns	CASES	NON-CASES	STATISTICS	Pg. # & NOTES
		Questionnaire Admin data Other (specify) Unclear	Intervention	Intervention	Intervention	RR (risk ratio) OR (odds ratio) SE (standard error) 95% CI DF P- value (enter exact p value if available) Chi2 Other Covariates (control variables, age, gender, education, ethnicity, duration dependence, labour market conditions, other)	
			Comparison	Comparison	Comparison		

Repeat as needed

OUT COME DATA

TIME-TO-EVENT OUTCOME DATA

OUTCOME	TIME POINT (s) (record exact time from beginning of unemployment, there may be more than one, record them all)	SOURCE	Method of estimation			STATISTICS	Pg. # & NOTES
		Questionnaire Admin data Other (specify) Unclear	Non-parametric Semi-parametric Parametric			HR (hazard ratio) SE (standard error) 95% CI DF P- value (enter exact p value if available) Chi2 Other Covariates (control variables, age, gender, education, ethnicity, duration dependence, labour market conditions, other)	

Repeat as needed

CONTINUOUS OUTCOME DATA

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

*Repeat as need

1.3 ASSESSMENT OF RISK OF BIAS

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. <i>A priori</i> protocol? ^d		
9. <i>A priori</i> 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.¹⁴ 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, 2008). 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)

¹⁴ 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).

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
 - 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 <i>identifying</i> relevant confounders described by researchers:	yes no	<input type="checkbox"/> <input type="checkbox"/>
If yes, describe the method used:		
Relevant confounders described:	yes no	<input type="checkbox"/> <input type="checkbox"/>
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	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Age	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Education	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ethnicity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Labour market conditions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unemployment duration	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Censoring	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Unobservables ¹⁵	<input type="checkbox"/>	Irrelevant	<input type="checkbox"/>	<input type="checkbox"/>
Other:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

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 behavior 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.

¹⁵ See user guide for unobservables

1.4 SEARCH STRATEGY

1. DE "Social Security"
2. DE "WELFARE recipients"
3. welfare w1 payment*
4. welfare w1 recipient*
5. welfare w1 support*
6. economic w1 support*
7. public w1 assistance*
8. welfare w1 payment*
9. public w1 support*
10. financial w1 support*
11. welfare w1 service*
12. direct* w1 payment*
13. general w1 assistance
14. Social w1 Support
15. cash w1 assistance
16. income w1 assistance
17. benefit*
18. social w1 assistance*
19. social w1 securit*
20. social w1 welfare
21. social w1 allowance*
22. insurance w1 benefit*
23. social w1 benefit*
24. welfare w1 benefit*
25. TANF
26. Insurance*
27. 1-26/or
28. DE "EMPLOYABILITY"
29. Employ*
30. Job*
31. work*
32. un-employ* or unemploy*
33. re-employ* or reemploy*
34. 28-33/or
35. effect*
36. threat*
37. incentive*
38. disincentive*

39. impact*
40. motivat*
41. 35-40/or
42. Expir*
43. Lapse*
44. Terminat*
45. Duration*
46. Generosit*
47. Change*
48. Entitl*
49. Length
50. Extend*
51. Extension*
52. Exhaust*
53. exit*
54. 42-53/or
55. 27 and 34 and 41 and 54