Research on Social Work Practice
1-29

DOI: 10.1177/1049731516630385
rsw.sagepub.com
SAGE

Active Labour Market Programme Participation for Unemployment Insurance Recipients: A Systematic Review

Trine Filges[1], Geir Smedslund[2], and Anne-Marie Klint Jørgensen[1]

Abstract

Objective: This review evaluates the effectiveness of Active Labour Market Programme (ALMP) participation on employment status for unemployment insurance recipients. **Methods and Analysis:** We followed Campbell Collaboration guidelines to conduct a systematic review. **Results:** A total of 73 studies met the inclusion criteria and were critically appraised by the review authors. The available evidence suggests that there is a general effect of participating in ALMP. The findings are mixed, however, depending on the approach used to investigate the effect, with no effect found of being assigned to ALMP participation at a particular *moment*. **Authors' Conclusions:** The available evidence does suggest that there is an effect of participating in ALMP, but the effect is small and we found no effect of being assigned to ALMP participation at a particular moment.

Keywords

unemployment, field of practice, meta-analysis, outcome study, quantitative, outcome study

Introduction

During the 1990s, many countries introduced Active Labour Market Programmes (ALMPs) in an effort to reduce unemployment. Public spending on labor market programs is typically split into the so-called active and passive measures (Martin, 2000). In 2012, the average spending on active measures across the Organization for Economic Cooperation and Development (OECD) countries was 0.6% of gross domestic product (GDP) and 0.9% of GDP was spent on passive measures (OECD Database on Labour Market Programmes; www.oecd.org/employment/database). The active measures comprise a wide range of policies aimed at improving the access of the unemployed to the labor market and jobs, while the passive measures relate to spending on income transfers, protecting individuals against loss of income and providing unemployed individuals the possibility of finding a better match between their qualifications and job vacancies. (Filges, Geerdsen, Knudsen, & Jørgensen, 2014). In countries such as Australia, United States, Denmark, Sweden, England, and Switzerland, participation in an ALMP is required if an unemployed individual is to continue receiving benefits (Geerdsen, 2003; Gerfin & Lecher, 2002). Typically, compulsory program participation is required after the individual has received unemployment benefits for a certain period of time.

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

We focus on research on the outcome of program participation, that is, effects *during* and *after* program participation (Heckman, Lalonde, & Smith, 1999; Martin & Grubb, 2001). The effects of ALMP participation on job-finding rates are typically composed of two separate effects: a lock-in effect and a post-program effect. The lock-in effect refers to the period of participation in a program. 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 ongoing skill-enhancing activity; hence the lock-in effect. The post-program effect refers to the period after participation in a program. If the ALMP has increased the individual's employability, a rise in the job-finding rate is expected. The combination of these two effects consequently determines the net effects of ALMP participation on unemployment duration.

[1] The National Institute of Social research, Copenhagen, Denmark
[2] Norwegian Institute of Public Health, Oslo, Norway

Corresponding Author:
Trine Filges, The National Institute of Social research, Herluf Trolles Gade 11, Copenhagen K, 1052, Denmark.
Email: tif@sfi.dk

Description of the Intervention

In this review, the intervention is ALMP participation by those in receipt of unemployment insurance benefits. However, studies in which the participants are a mix of individuals receiving unemployment insurance benefits and individuals receiving other types of unemployment benefits are included if more than 60% of the participants receive unemployment insurance benefits. A large variety of different ALMPs exist among countries. They can consist of job search assistance, training, education, subsidized work, and similar programs. Some of the programs (e.g., subsidized work, training, and education) demand full-time participation over a long time period (e.g., several months), while other programs (e.g., job search assistance and education) are part-time and have a short duration (e.g., few days/weeks). It is possible to classify these programs into a set of four core categories: A, (labor market) training; B, private sector programs; C, direct employment programs in the public sector; and D, job search assistance. The categories we use broadly correspond to classifications that have been suggested and used by the OECD and Eurostat (Eurostat, 2005; OECD, 2004), even though there are differences between OECD and Eurostat in how they define and categorize these programs. The four categories are described below in detail:

A. The first program type, (*labor market*) *training*, encompasses measures such as classroom training, on-the-job training, and work experience. The training can either provide a more general education (as with language courses or basic computer courses) or specific vocational skills (as with advanced computer courses or courses providing technical or manufacturing skills). Their main objective is to develop the productivity and employability of the participants and to enhance human capital by increasing skills. Training programs constitute the ''classic'' component of ALMP.
B. *Private sector programs* are those aimed at creating incentives to alter employer and/or worker behavior in relation to private sector employment. Wage subsidies are the most commonly used measure in this category. The objective of subsidies is to encourage employers to hire new workers or to maintain jobs that would otherwise be broken up. These can either be direct wage subsidies to employers or financial incentives that are offered to workers for a limited period of time. The use of self-employment grants form another type of subsidized private sector employment: These grants may be offered to participants who start their own business, sometimes along with advisory support for a fixed period of time (Eurostat, 2005; OECD, 2004).
C. In contrast to subsidies in the private sector, the third program type, *direct employment programs in the public sector*, focuses on the direct creation and provision of public works or other activities that produce public goods or services. These measures are mainly targeted at the most disadvantaged individuals, pursuing the aim of keeping them in contact with the labor market and precluding the loss of human capital during a period of unemployment. The created jobs are, nevertheless, often additionally generated and at a distance from the ordinary labor market.
D. The fourth type of program, *job search assistance*, encompasses measures aimed at enhancing job search efficiency. The services included are job search courses and related forms of intensified counselling for those who have difficulty finding employment. The public employment services often target the disadvantaged and long-term unemployed, whereas private services may focus on the more privileged employees and white-collar workers. These programs are usually the least expensive.

How the Intervention Might Work

ALMPs were adopted by most advanced countries during the 1990s (Gerfin & Lechner, 2002). The declared purpose of such policies is to protect workers who are exposed to negative employment shocks due to changing market conditions (Aarnio, 1996; Filges, Kennes, Larsen, & Tranæs, 2011).

Programs that involve subsidized work, training, and education are designed to reduce skill loss during extended periods of unemployment and to redirect the skills of those who are left without work as a result of new technology or increased international trade (Kluve et al., 2007). The introduction of ALMPs is thus often motivated by the need to upgrade the skills of especially those suffering long-term unemployment to improve their productivity and, subsequently, their employability. If participation in an ALMP increases the individual's employability, a rise in the job-finding rate is to be expected; however, the increased human capital may result in higher reservation wages (the minimum wage at which a job offer is acceptable), effectively offsetting the positive employment effect (Filges et al., 2011; Mortensen, 1987). Moreover, some programs may stigmatize workers in the view of potential employers. Programs associated with participants having poor employment prospects (e.g., the long-term unemployed) may carry a stigma. Because of asymmetric information (a situation where there is imperfect knowledge where one party has different information from another), employers cannot know the productivity of new workers, some of whom they might hire from the pool of the unemployed. Prospective employers might then perceive participants in such employment programs as low-productivity workers or as workers with a tenuous labor market attachment (Kluve, Lehmann, & Schmidt, 1999; Kluve et al., 2007).

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

Why It Is Important to do This Review

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

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

The effect of ALMPs for unemployed people receiving other kinds of unemployment benefits is reviewed in the Campbell Systematic Review ''Work programmes for welfare recipients'' (Smedslund et al., 2006) where the objective was to estimate the effects of work programs on welfare recipients' employment and economic self-sufficiency. Individuals who are entitled to unemployment insurance benefits or who have pensions of any kind were, however, excluded in Smedslund et al. (2006).

To the best of our knowledge, there is currently no systematic review on the effect of ALMP participation in unemployed individuals receiving unemployment insurance benefits—the focus of this review.

Objective of the Review

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

Methods

Systematic review methods, following the Campbell Collaboration (2014) guidelines, were used to conduct this study and meta-analytic methods were used to synthesize study results. The protocol for this review is registered and published in the Campbell Collaboration library (Filges, Geerdsen, Smedslund, Knudsen, & Jørgensen, 2013).

Criteria for Considering Studies for This Review

Types of studies. The study designs eligible for inclusion in the review were controlled trials, randomized controlled trials (RCTs), quasi-randomized controlled trials (QRCTs), where participants are allocated by, for example, alternate allocation, participant's birth date, date, case number, or alphabetically, and nonrandomized controlled trials (NRCTs) where participants are allocated by other actions controlled by the researcher. Nonrandomized studies (NRSs) where allocation is not controlled by the researcher and two or more groups of participants are compared. Participants are allocated by, for example, time differences, location differences, decision makers, policy rules, or participant preferences.

All study designs that used a well-defined control group were eligible for inclusion in this review. Studies that utilized qualitative approaches were not included due to the absence of adequate control group conditions. We only included studies that used individual microdata. We excluded studies that rely on regional or national time-series data, even though microeconometric estimates of individual treatment effects merely provide partial information about the full impact of ALMP.

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

Types of participants. The participants were required to be unemployed individuals who received unemployment insurance benefits. The International Labour Office (ILO) definition of an unemployed individual is a person, male or female, aged 15–74, without a job who is available for work and either has searched for work in the past 4 weeks or is available to start work within 2 weeks and/or is waiting to start a job already obtained; however, different countries may apply different definitions of an unemployed individual, see for example Statistics Denmark (2009). The eligibility rules of unemployment insurance benefits differ between countries. We excluded individuals receiving other types of benefits such as social assistance benefits or benefits not related to being unemployed. Studies including a mix of individuals receiving unemployment insurance benefits and other individuals receiving social assistance benefits and/or other types of benefits were only included if more than 60% of the included individuals received unemployment insurance benefits.

Types of interventions. The intervention is participation in ALMP. ALMPs can include a wide range of activities as listed below. ALMPs typically apply to unemployment insurance beneficiaries and (if different) employable social assistance beneficiaries, but similar principles are increasingly being applied to lone-parent and disability beneficiaries. In this

review, ALMPs were understood in the narrow sense of training or employment measures for the unemployed receiving unemployment insurance benefits.

A large variety of different ALMP programs exists among countries that may be classified into four core categories. In this review, we adopted categories that broadly correspond to classifications suggested and used by the OECD and Eurostat (Eurostat, 2005; OECD, 2004), even though there are differences between OECD and Eurostat in how they define and categorize these programs. The four categories are as follows: A, (labor market) training; B, private sector programs; C, direct employment programs in the public sector; and D, job search assistance (see previous section Description of the Intervention for more details).

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

Types of outcomes. The objective of this review was to study the effect of ALMP participation on employment status. Our main interest was to include studies in a meta-analysis where hazard ratios (HRs) and variance were either reported or calculable from the available data. The HR measures the proportional change in hazard rates (defined as the event rate [finding a job] at time t conditional on survival [staying unemployed] until time t or later) between unemployed persons who have participated in ALMPs and unemployed persons who have not participated in ALMPs. The primary outcome was exits from the unemployment insurance system and into employment. Studies that only examine exits to other destinations, such as other types of social benefits or nonemployment, were not included. The included studies reported outcomes in the form of HRs and risk difference (the difference in the probability of employment) or data that permitted the calculation of an HR or risk difference.

In addition to the primary outcome, we considered secondary outcomes that are relevant to the impact ALMP has on the duration of employment and on income.

A few studies provided data on the exit rate from reemployment. We included the measure of exit rate from reemployment in the analysis of secondary outcomes. A higher exit rate from reemployment may indicate that the participation in ALMP forces unemployed individuals to find jobs that do not match their qualifications and, therefore, to return to unemployment quickly.

Primary outcomes: (a) relative exit rate from unemployment to employment (measured as HR) and (b) difference in probability of employment (measured as risk difference).

Secondary outcome measures: (a) duration of first employment spell postintervention, (b) relative exit rate from reemployment to unemployment (measured as HR), and (c) reemployment income.

Search Methods for Identification of Studies

The search was performed by one review author (A.K.J.) and one member of the review team (Pia Vang Hansen [P.V.H.]). Members of the review team at SFI Campbell were the research assistants PVH, Simon Helth Filges (SHF), and Trondur Møller Sandoy (TMS). Relevant studies were identified through electronic searches of bibliographic databases, government policy databanks, Internet search engines, and hand searching of core journals. We searched to identify both published and unpublished literature. The searches were international in scope. Reference lists of included studies and relevant reviews were also searched. For additional details of the search methods, see Filges, Smedslund, Knudsen, and Jørgensen (2015).

Selection of Studies

One review author (Anne-Sofie Due Knudsen [A.D.K.]) and two members of the review team (S.H.F. and T.M.S.) independently read titles and available abstracts of reports and articles identified in the search to exclude reports that were clearly irrelevant. Citations considered relevant by at least one reviewer were retrieved in full-text versions. If there was insufficient information in the title and abstract to judge relevance, the full text was retrieved.

Two reviewers (A.D.K. and T.F.) and two members of the review team (S.H.F. and T.M.S.) read the full-text versions to ascertain eligibility based on the selection criteria. Any disagreements were resolved by discussion. A screening guide (see Filges et al., 2015) was used to determine inclusion or exclusion and was provided in the protocol (Filges et al., 2013).

Data Extraction and Management

One review author (A.D.K.) and two members of the review team (SHF and TMS) independently coded the included studies (see Filges et al., 2015). A coding sheet was piloted on several studies (Filges et al., 2013). Any disagreements were resolved by discussion. Information was extracted on characteristics of participants, intervention characteristics and control conditions, research design, sample size, and censoring. Numeric data extraction (outcome data) was performed by one review author (A.D.K.) and checked by a second review author (T.F.). Extracted data were stored electronically. Analysis was conducted in RevMan5 Version 5.1 and STATA Version 14.

Assessment of Risk of Bias in Included Studies

Two review authors (T.F. and A.D.K.) independently assessed the risk of bias for each included study. There were only minor disagreements and these were resolved by discussion. We assessed the methodological quality of studies using a risk of bias model developed by Professor Barnaby Reeves in association with the Cochrane Non-Randomised Studies Methods Group. This risk of bias model was introduced by Prof. Reeves at a workshop on risk of bias in NRS at SFI Campbell, February 2011. The model is a further development of the work carried

out in the Cochrane Non-Randomised Studies Method Group (see Filges et al., 2015). This model, an extension of the Cochrane Collaboration's risk of bias tool, covers risk of bias for RCTs as well as risk of bias for NRS that have well-defined control groups. The extended model is organized and follows the same steps, as the risk of bias model described in the Cochrane Handbook, Chapter 8 (Higgins & Green, 2011).

Risk of bias judgment items. The risk of bias model used in this review is based on 9 items. The 9 items refer to sequence generation, allocation concealment, confounders, blinding, incomplete outcome data, selective outcome reporting, other potential threats to validity, a priori protocol, and a priory analysis plan.

Confounding

An important part of the risk of bias assessment of NRS (NRCT and NRS) is consideration of how the studies deal with confounding factors (see Filges et al., 2015). Selection bias is understood as systematic baseline differences between groups that can therefore compromise comparability between groups. Baseline differences can be observable to the researcher (e.g., age and gender) and unobservable (e.g., motivation and ''ability''). There is no single NRS design that always solves the selection problem. Different designs attempt to provide solutions to the problem of potential selection bias under different assumptions and consequently require different types of data. Designs particularly vary with respect to how they deal with selection on ''unobservable'' factors. The ''right'' method depends on the model generating participation, that is, assumptions about the nature of the process by which participants are selected into a program.

As there is no universal correct way to construct counterfactuals for nonrandomized designs, we looked for evidence that identification was achieved and that the authors of the primary studies justified 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 participating in ALMP and an individual not participating in ALMP is exactly the participation and that the source of difference between their participation status is not endogenous to the individuals' exit rate to employment. The judgment is reflected in the assessment of the confounder ''unobservables'' in the list of confounders considered important at the outset and defined in the protocol for this review.

In addition to unobservables for this review, we identified the following observable confounding factors to be the most relevant: age, gender, education, ethnicity, labor market conditions, censoring, and unemployment duration. In each study, we assessed whether these factors had been considered, and in addition, we assessed other factors likely to be a source of confounding within the individual included studies.

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

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

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

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

Measures of treatment effect. The treatment effect was measured either as the impact on the hazard rate or as the impact on the probability of employment at some date or time interval after the completion of the program. Our main interest was to include studies in a meta-analysis where HRs and variances were either reported or were calculable from the available data.

The HR measures the proportional change in hazard rates between unemployed individuals who have participated in ALMPs and unemployed individuals who have not participated in ALMPs. 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.

The majority of studies provided HRs and variances or data enabling the calculation of HRs and variances. The acceptable outcome measurement frequency for calculating HRs in this review was 3 months or less. A study reporting only outcomes measured on time intervals of more than 3 months was not included in the meta-analysis.

As stated in the protocol, Filges et al. (2013), individual participant data were *not* requested to calculate log HRs, as this may introduce bias due to the time span of studies (the time span between the earliest we knew of and the latest is 30 years).

Some studies reported risk difference and variances or data that enabled the calculation of risk difference and variance. The risk difference is the difference in the probability of employment at a given moment or in a given time period.

If risk difference and variances were not reported, they were computed directly using the observed number of events and the

total number of participants (Borenstein, Hedges, Higgins, & Rothstein, 2009).

We separately pooled studies where outcomes were measured (or could be calculated) as HRs and risk difference. We performed the meta-analyses using the log HR and variance and the risk difference and variance. We report the 95% CIs.

The secondary outcomes were also measured as HRs and the effect sizes as log HRs by two studies and in addition one study provided data on earnings that permitted the calculation of an effect size (Hedges' g was used for estimating standardized mean differences [SMDs]) and two studies reported the effect on the duration of reemployment (calculation of an SMD was not possible but both studies reported the mean difference measured in months with variances). The different outcomes were analyzed separately and we report the 95% CIs.

Further, we analyzed the effects measured by HRs obtained using the so-called timing-of-events approach separately from effects measured by HRs obtained using other methods. These other methods used in the included studies are randomized assignment, matching, instrument variables, and multiple regression.

The timing-of-events approach is special, as it explores information on the *timing* of events (like the moment when the individual enrols in training and the moment he finds a job) to estimate the individual training effect. The training effect obtained using this approach is the effect on the exit rate to work of being assigned to training at a particular moment as opposed to the effect of being assigned to training in general. The empirical approach involves estimation of models simultaneously explaining the duration of unemployment before obtaining work or participating in training programs.

For individuals who enter training at time t, the natural control group consists of individuals unemployed for the same period of time at t, but who have not yet received training. A necessary condition for identification of an effect is that there is *some* randomization in the training assignment at that particular t. The model allows for selection effects by way of unobserved determinants that affect the treatment assignment as well as the outcome. It is thus not necessary to make a conditional independence assumption, that is, all determinants of the process of treatment assignment are captured by the data (the covariates used in the model) so that the remaining variation in assignment to treatment is independent of the determinants of the outcome. The timing-of-event model framework allows for randomization because it specifies assignment by the *rate* of entering training. Thus, there is a random component in assignment in a small time interval that is independent of the covariates. An essential assumption when using the timing-of-events approach is thus the no anticipation assumption. Individuals may know the determinants of the process leading to training, including the probability distribution of the duration until training, but it is assumed that they do not know the outcome of this process, the realization of the moment of assignment, in advance. The random *realization of the exact moment of assignment* is what identifies the effect and the effect obtained is the *effect of treatment at time t*.

Unit of analysis issues. To account for possible statistical dependencies, we examined a number of issues: whether individuals were randomized in groups (i.e., cluster randomized trials), whether individuals had undergone multiple interventions, whether there were multiple treatment groups, and whether several studies were based on the same data source.

Cluster randomization. No studies using cluster randomization were found.

Multiple intervention groups. Two studies, analyzing ALMP in Germany, provided results separately for East and West Germany. We used the effect estimates from East and West Germany separately in the meta-analysis. Further, one study provided the results of participating in ALMP in West Germany for the years 1986 and 1993 separately. We used the effect estimates from 1986 and 1993 separately in the meta-analysis. Finally, one study analyzed an RCT conducted in Florida and Washington, DC. Results were reported separately for the two states, and we used the effect estimates separately in the meta-analysis.

Where studies reported separate effect estimates, for example, separated by gender or by ALMP type, a synthetic (average) effect size was calculated and used in the analyses of the overall effect of ALMP participation to avoid dependence problems.

Multiple interventions per individual. There were no studies with multiple interventions per individual.

Multiple studies using the same sample of data. Three studies used the same sample of data, that is, the studies used administrative register data from the same country covering the same time period. All three studies used data from Switzerland where the administrative registers provide complete coverage; that is, all registered unemployed in the selected period are included in the administrative registers. Complete coverage of administrative registers applies to other countries as well. Two primary studies analyzed a random sample from these administrative registers in Switzerland covering ALMP participation in 1998 and one primary study analyzed a complete sample from one canton covering ALMP participation in 1998. The data used in these primary studies were thus (partly) representative of the same population of unemployed at the same time. We reviewed all three studies, but in the meta-analysis we only included one estimate of the ALMP participation effect from this sample of data. The choice of which estimate to include was based on our quality assessment of the studies. We chose the estimate from the study that we judged to have the lowest risk of bias paying particular attention to the confounding item. Two studies had equal scoring on the confounding item and we based the choice on the incomplete data item and sample selection choices.

Multiple time points. All studies either reporting HRs or where calculation of HRs were possible reported the effect from the end of treatment. For the studies reporting the effect of timing

of the event (participation in ALMP), all studies reported the effect on the hazard rate from end of treatment and some studies in addition reported the effect on the hazard rate from the beginning of treatment. One study reported the effect from the beginning of treatment only. Each time point, start of treatment and end of treatment, was analyzed in a separate analysis. For the studies reporting risk difference (or where it was possible to calculate risk difference), it was possible to pool all the studies and we used the outcome measured closest to 1 year after treatment.

Dealing with missing data and incomplete data. Missing data and censoring were assessed in the included studies. For studies using questionnaire data, a sensitivity analysis was performed to assess potential bias. For studies, using time to event data in which the censoring level was high (more than 25%) or the level was not reported, a sensitivity analysis was performed to assess potential bias in the analysis. Attrition rates, reasons for attrition and whether intention-to-treat analysis (ITT) was conducted were recorded where possible from included RCTs and QRCTs. It was not possible to perform a sensitivity analysis as all RCTs and QRCTs conducted ITT analysis.

Assessment of heterogeneity. Heterogeneity among primary outcome studies was assessed with χ^2 (Q) test and the I^2 and τ^2 statistics (Higgins, Thompson, Deeks, & Altman, 2003). Any interpretation of the χ^2 test was made cautiously on account of its low statistical power.

Assessment of reporting bias. We used funnel plots to identify possible publication bias.

Grading of evidence. The quality of evidence was assessed according to a systematic and explicit method (Guyatt et al., 2008). In order to indicate the extent to which one can be confident that an estimate of effect is correct, judgments about the quality of evidence were made for each comparison and outcome. These judgments considered study design (RCTs, QRCTs, NRCTs, and NRSs), study quality (detailed study design and execution), consistency of results (similarity of estimates of effect across studies), and directness (the extent to which people, interventions, and outcome measures were similar to those of interest). The following definitions were used in grading the quality of evidence (Balshem et al., 2011): *High*: We are very confident that the true effect lies close to that of the estimate of the effect. *Moderate*: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. *Low*: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect. *Very low*: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect. Any estimate of effect is very uncertain.

Data Synthesis

As planned (outlined in the protocol, Filges et al., 2013), we used random effects models to estimate the overall effect as ALMPs vary in their content and deal with diverse populations of participants and labor market conditions. Analysis was conducted in RevMan5, except the meta-regression which was conducted in STATA. Studies that were coded with a very high risk of bias (scored 5 on the risk of bias scale) were not included in the data synthesis.

As outlined in the section Units of Analysis Issues, it was possible to group outcomes as follows: HRs from end of treatment and risk difference approximately 1 year after treatment as possible. As lock-in effects may be considerable, effects where lock-in effects were considered were analyzed separately from estimates of post-program effects. The lock-in effect refers to the period of participation in a program. Only studies using the timing-of-events method measured results where lock-in effects were considered. The proportional shift of the hazard rate was measured from the beginning of treatment, thus combining the lock-in and post-program effects. The combination of these two effects consequently determines the net effects (net of lock-in) of ALMP participation on the exit rate to employment.

We planned to distinguish between the counterfactual situations. The main distinction between counterfactual situations is whether the studies estimate an effect relative to a control group who is never going to participate or they estimate an effect relative to a control group who may participate at a later point in time. Only two studies estimated an effect relative to a control group who was never going to participate so we did not distinguish between the counterfactual situations.

When the effect sizes used in the data synthesis were HRs, they underwent log transformations before being analyzed. 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 an HR 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. Graphical displays for meta-analysis performed on ratio scales sometimes use a log scale, as the CIs then appear symmetric. This is however not the case for the software RevMan5 used in this review. The graphical displays use HRs and the mean effect size is reported as an HR.

All analyses were 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 were calculated using 95% CIs.

Moderator analysis and investigation of heterogeneity. With the aim of explaining observed heterogeneity, we planned to investigate the following factors: type of ALMP (labor market training, private sector programs, direct employment programs in the public sector, and job search assistance), study-level

summaries of participant characteristics (e.g., studies considering a specific age group, gender, or educational level or studies where separate effects for men/women, young/old, or low/high educational level are available), and labor market conditions.

It was not possible, however, to investigate the impact of either participant characteristics or labor market conditions. Among the studies used in the data synthesis, only three restricted its analysis to a specific age-group and none restricted their analyses to a specific educational level. No separate estimates within studies for young/old or low/high educational levels were available. Seven studies provided separate effect estimates by gender. Three of these reported risk difference and variances or data that enabled the calculation of risk difference and variance. Four studies reported HRs and variances or data that enabled the calculation of HR and variance. One of these used the timing-of-events approach. The majority of studies did not report the labor market conditions and there was hardly any variation in this covariate among those that did.

It was possible to undertake a moderator analysis of different types of ALMP in order to explore potential differences in effects for the following outcomes:

- Risk difference post participation,
- HR net of lock in using the timing-of-events approach, and
- HR post participation using the timing-of-events approach.

In summary, it was possible to analyze only one moderator ("type of ALMP") of the five moderators we had planned to investigate (Filges et al., 2013), and then only for the outcomes mentioned above. Several of the included studies provided results separated by type of ALMP. We performed single-factor subgroup analysis. The subgroup analyses were 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 for each subgroup were calculated using 95% CIs.

The assessment of any difference between subgroups was based on 95% CIs. No conclusions from single-factor subgroup analyses were drawn and interpretation of relationships was cautious, as they were based on subdivision of studies and indirect comparisons.

In addition to the risk difference post participation, outcome was investigated using meta-regression. Conventional meta-regression techniques rely on the assumption that effect size estimates from different studies are independent and have sampling distributions with known conditional variances. The independence assumption is violated when studies produce several estimates based on the same individuals who are the case in the present context where studies report effect sizes separated by ALMP type; the model was therefore estimated using the robust standard error method (Hedges, Tipton, & Johnson, 2010). This more robust technique is beneficial because it takes into account the possible correlation between effect sizes separated by ALMP type within the same study and allows all of the effect size estimates to be included in meta-regression.

Since this robust standard error method uses degrees of freedom based on the number of studies (rather than the total number of effect sizes), it was only possible to perform an analysis for risk difference post participation (17 studies were included in the analysis). For the remaining outcomes, there were insufficient studies to perform a meta-regression using the robust standard error method. The technique used 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 CIs and p values generated this way typically reflect the correct size in samples, requiring between 20 and 40 studies.

An important feature of this more robust standard error analysis is that the results are valid regardless of the weights used. For efficiency purposes, we calculated the weights using a method proposed by Hedges, Tipton, and Johnson (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. For the results we calculated, weights were used based on estimates of τ^2, where $\rho = .80$. Sensitivity tests were also conducted using a variety of ρ values; these indicated that the general results and estimates of the heterogeneity were robust to the choice of ρ. The residual variance component was estimated using the method-of-moments estimator.

Conclusions from meta-regression analysis were drawn with caution and were not based on significance tests.

Sensitivity Analysis

Sensitivity analysis was used to evaluate whether the pooled effect sizes were robust across study design (RCT, QRCT, and NRS) and components of the risk of bias tool. For risk of bias, we performed sensitivity analysis for the sequence generation (only for RCTs and QRCTs), confounding (only for NRSs), incomplete data, and selective reporting items of the risk of bias checklists, respectively. Sensitivity analysis was further used to examine the robustness of conclusions in relation to the quality of data (outcome measures based on weekly, monthly, or quarterly data and whether data were based on questionnaires or administrative registers). The extent to which the results, measured by HRs, might be biased by a high censoring level was also included in the sensitivity analysis.

Results

We ran the searches during September 2012. The total number of potential relevant studies constituted 16,422 (gray literature

Table 1. Number of Included Studies.

Country	Total	Provide Effect Estimate[a]	Reduction Due to		Used in Data Synthesis
			Too High Risk of Bias	Overlap of Data Samples[b]	
Germany	15	11	0	0	11
United States	11	9	3	0	6
Sweden	10	5	1	0	4
Denmark	7	7	0	0	7
Switzerland	6	4	0	3	2
Norway	4	1	1	0	0
Israel	3	1	0	0	1
France	3	2	0	0	2
Austria	3	1	0	0	1
The Netherlands	2	2	0	0	2
United Kingdom	2	2	0	0	2
Finland	2	0	0	0	0
Hungary	2	0	0	0	0
Belgium	2	1	1	0	0
Romania	1	1	0	0	1
Total	73	47	6	3	39

Note. The reduction due to too high risk of bias preceded the reduction due to overlap of data sample.
[a]Or data that enable the calculation of an effect estimate. [b]The data samples used are representative of the same population at a given time (see *Multiple studies using the same sample of data* for this methodological issue).

search: 720; database search: 18,129; and snowball search: 18; hand-search of two journals; see Filges et al., 2015).

In total, 677 hits were retrieved for full-text screening. Of these, 534 did not fulfill the screening requirements and were excluded. No papers from hand searching or from the search of the gray literature were included. See the following section Excluded Studies for further details regarding excluded and unobtainable studies.

A total of 73 studies, consisting of 143 papers, met the inclusion criteria and were appraised by the review authors. Thirty-nine studies were included in the data synthesis. See Filges et al. (2015) for flowchart for the literature search and screening.

Description of the Studies

Studies included in the systematic review. The search resulted in a final selection of 73 studies that met the inclusion criteria for this review. Of these, 26 did not provide data sufficient for the calculation of an effect size (see Filges et al., 2015). In general, standard errors were not reported and no other information making it possible to calculate standard errors was reported (with a few exceptions where not even an effect size could be extracted). Of the remaining 47 studies, 6 were coded with a very high risk of bias (5 on the risk of bias scale) and were therefore not used in the data synthesis. Two studies could not be used in the data synthesis due to overlapping samples (i.e., three studies used administrative register data from the same country covering the same time period; see the Sensitivity Analysis section for discussion of this methodological issue). These studies analyzed ALMP in Switzerland. After these reductions, 39 studies remained and were included in the data synthesis (see Table 1 for included studies).

For studies with overlapping samples, the choice of which study to use in the data synthesis was based on our risk of bias assessments. The citations for the 47 studies that provided effect size estimates can be found in the Studies With Effect Estimate section.

A decision had to be made about which of the three studies from Switzerland (Frölich & Lechner, 2010; Gerfin, 2002; Prey, 2000), which used data (partly) representative of the same population of unemployed at the same time, to use in the data synthesis. One of the studies was deselected, as it was judged to have a higher risk of bias than the others due to confounding (Prey, 2000). The remaining two studies were judged to have the same risk of bias due to confounding as well as incomplete data, and the choice of which to include in the meta-analysis was based on the different sample selection choices of the two studies. The study with the least restrictive sample selection choice was included in the meta-analysis (Gerfin, 2002).

In total, the 39 studies provided effect estimates for 43 unique populations. Two studies (Caliendo, 2011; Hujer, 2010), analyzing ALMP in Germany, provided results separately for East and West Germany. We used the effect estimates from East and West Germany separately in the meta-analysis. Further, one study provided the results of participating in ALMP in West Germany for the years 1986 and 1993 separately (Völter, 2007). We used the effect estimates from 1986 and 1993 separately in the meta-analysis. Finally, one study analyzed an RCT conducted in Florida and Washington, DC (Decker, 2000). Results were reported separately for the two states and we used the effect estimates separately in the meta-analysis (see Filges et al., 2015, for details).

Table 2. Characteristics of Effect Estimates Used in the Data Synthesis.

Characteristics			
Country		Sample Size	
European countries (including Israel and excluding the Nordic countries)	25	Range	88–1,438,156
United States	7	Average	88,772
Nordic countries	11	Total	3,817,210
Analysis period		Type of data	
1980s	8	Administrative registers	31
1990s	21	Questionnaire	7
2000s	14	Combination	5
Effect estimates separated into two or more types of ALMP		Compulsory activation a part of the system	
Two types	10	Yes	37
Three types	3	No	2
Four types	3	Unclear	4
Types of ALMP effect estimates reported		Considered specific age group or education level	
Labor market training	17	Specific age group	3
Private sector programs	12		
Public employment programs	11		
Job search assistance	25		
Employment programs not separable into private/public	4	Specific education level	0
Unclear	2		
Considered specific gender or separated by gender		Labor market conditions	
Considered only females	1	Reported unemployment percent	8
Separated estimates by gender	7	Not reported	35

The characteristics of the 43 effect estimates that were used in the data synthesis are shown in Table 2 (see Filges et al., 2015, for a description of the individual studies).

The effect estimates provided were mainly from European countries. Seven effect estimates were from the United States and 11 were from the Nordic countries. Twenty-one effect estimates were provided using data from the 1990s. Eight effect estimates were provided using data from the 1980s and 14 effect estimates were provided using data from the period 2000–2008. Data were drawn mainly from administrative registers. The sample sizes were generally large; all but seven of the effect estimates provided were obtained using sample sizes of more than 1,000. The majority of effect estimates were obtained in circumstances where the counterfactual situation was potential participation in ALMP at a later time; only two studies providing two effect estimates reported that the counterfactual situation was no participation. The analyses were restricted to a specific age-group (either young, 25 years old or younger; or old, aged 55 or 56 years) in only three cases, and none were restricted to analyze a specific educational level. One analysis included only females and seven effect estimates were provided separated by gender. The majority of studies did not report the labor market conditions (unemployment rate, vacancy rate, and/or labor market tightness [number of vacant jobs per unemployed])—only eight studies (providing eight effect estimates) reported labor market conditions in the form of the unemployment level.

More than half of the effect estimates provided were obtained from analyses including one type of ALMP only; 16 effect estimates were reported separated into two or more types of ALMP. The majority of programs could be classified into one of the four prespecified categories: (1) labor market training, (2) private sector programs, (3) public employment programs, and (4) job search assistance. Effect estimates of job search assistance and labor market training were those reported most frequently. Four of the programs (in Firth, 1999; Lalive, 2008; Caliendo, 2011, for both East and West Germany) could not be categorized as either private or public and two programs (in Frölich & Lechner, 2010b; Bennmarker, Nordström Skans, & Vikman, 2012b) could not be classified.

Excluded studies. In addition to the 73 studies that met the inclusion criteria for this review, several studies at first sight appeared relevant for the review but did not end up meeting our criteria. In 52 studies, the share of participants receiving unemployment insurance benefits was unclear/too low. Two studies used different data sources for treated/control and one study analyzed a program that could not be classified as ALMP. None of these studies fulfilled our inclusion criteria and were therefore not included in the final review (see the excluded studies and the reasons for exclusion in Filges et al., 2015).

Studies awaiting classification. Three studies could not be as they were written in German and, no one in the review team was able to read a German text. One reference was not obtained in full text, despite repeated attempts to locate it (see Filges et al., 2015, for details).

Risk of bias in included studies. Table 3 provides a summary of the risk of bias associated with the 47 studies from which it was possible to extract an effect. Of the included studies, 32 used nonrandomized designs and 15 studies used randomized designs. Of these, 13 were classified as RCTs and 2 as QRCTs. Only one of the studies had an a priori protocol and an a priori analysis plan.

Six studies were given a score of 5 (five studies on the confounding item and one study on the other bias item), corresponding to a risk of bias sufficiently high for the findings not to be considered in the data synthesis. For these six studies, we

Table 3. Risk of Bias—Distribution of the 47 Studies Reporting an Effect Size.

Risk of Bias Item	Judgment								Total Number of Studies
	High	Low	Unclear	1	2	3	4	5	
Sequence generation[a]	29	5	7	—	—	—	—	—	41
Allocation concealment[a]	29	3	9	—	—	—	—	—	41
Blinding[a]	—	—	1	0	0	34	6	0	41
Incomplete data[a]	—	—	2	19	6	8	6	0	41
Selective reporting[a]	—	—	1	30	4	3	3	0	41
Other bias[b]	—	—	0	40	1	0	0	1	42
Confounding[c]	—	—	0	15	5	2	4	5	31

Note. The judgment is based on a 5-point scale where 1 indicates low risk of bias and 5 indicates high risk of bias.
[a]Not judged for six of the studies as five studies scored 5 on the confounding item and one study scored 5 on the other bias item and was thereby not included in the data synthesis. Therefore, it was not relevant to judge on the remaining items for these six studies. [b]Not judged for the five studies scoring 5 on the confounding item. [c]Not judged for the 15 studies using randomized designs and neither for the one study that scored 5 on the other bias item.

did not judge the other items because they were already excluded from the data synthesis due to high risk of bias.

All studies using nonrandomized designs were judged to have a high risk of bias on the sequence generation item and the allocation concealment item. Of the 15 studies using randomized designs, only 5 were judged to have a low risk of bias on the sequence generation item and 3 on the allocation concealment item.

Due to the nature of the intervention, those in the treatment condition will always be aware that they are treated; therefore, assessment of the blinding item with regard to the participants did not differ across studies. The majority of studies were thus judged three on the blinding item. Furthermore, the nature of the outcome, employment, is objective and in the majority of studies data were obtained from administrative registers or questionnaires that were not collected with the aim of analyzing ALMP participation. Six studies departed from these considerations concerning the nature of the outcome, which is reflected in the score on their blinding item.

The majority of the included studies obtained data from administrative registers. For those studies using administrative registers providing complete coverage, very few were lost due to missing data and they scored 1 on the incomplete data item. Other studies used administrative registers not providing complete coverage of the outcome variables or used questionnaires subject to missing data. The extent of missing data is reflected in their score on the incomplete data item.

Selective reporting was judged not to be a concern in the majority of studies. In many studies, sensitivity analyses were conducted, with the results of these being reported together with very detailed information concerning the application of the estimation procedures used. Deviations are reflected in the score on the selective reporting item. The ''other bias'' item was not judged to be a concern, except for one study which scored 5 on this item.

Assessment of risk of bias due to confounding was not relevant for the 15 studies using randomized designs and neither for the one study that scored 5 on the other bias item. As mentioned, five studies scored 5 on the confounding item. The majority of the remaining studies were assessed to be of no concern regarding confounding or only of minor concern (see the detailed assessment of confounding in Filges et al., 2015).

Effects of the Intervention

In order to carry out a meta-analysis, every study must have a comparable type of effect size. The majority of studies reported HRs and variances (19) or provided data enabling the calculation of HRs and variances (5). The remaining studies reported risk difference and variances (13) or data that enabled the calculation of risk difference and variance (2).

For the studies reporting HRs and variances (or where it was possible to calculate HRs and variances) not using the timing-of-event approach, the length of the time periods after participation varied. The majority of studies (10) had data covering less than a year after participation and the remaining studies (4) had data that covered less than 2 years after participation. For the studies using the timing-of-event approach, the length of the time periods after participation/from beginning of treatment varied but on average studies had data covering 4 years after participation/from beginning of treatment.

For the studies reporting risk difference (or where it was possible to calculate risk difference), we used the outcome measured closest to 1 year after treatment. Seven studies reported outcome 1 year after participation, eight studies reported outcomes 1–2 years after participation, and three studies reported outcomes 4–5 years after participation (see Filges et al., 2015, for characteristics of the included studies).

The post effects measured by HRs and risk difference were pooled separately and studies using the timing-of-events approach were pooled separately.

Only studies using the timing-of-event approach reported effects net of locking-in. The proportional shift of the hazard rate was measured from the beginning of treatment, thus combining the lock-in (the lock-in effect refers to the period of participation in a program) and post-program effects. The combination of these two effects consequently determines the net

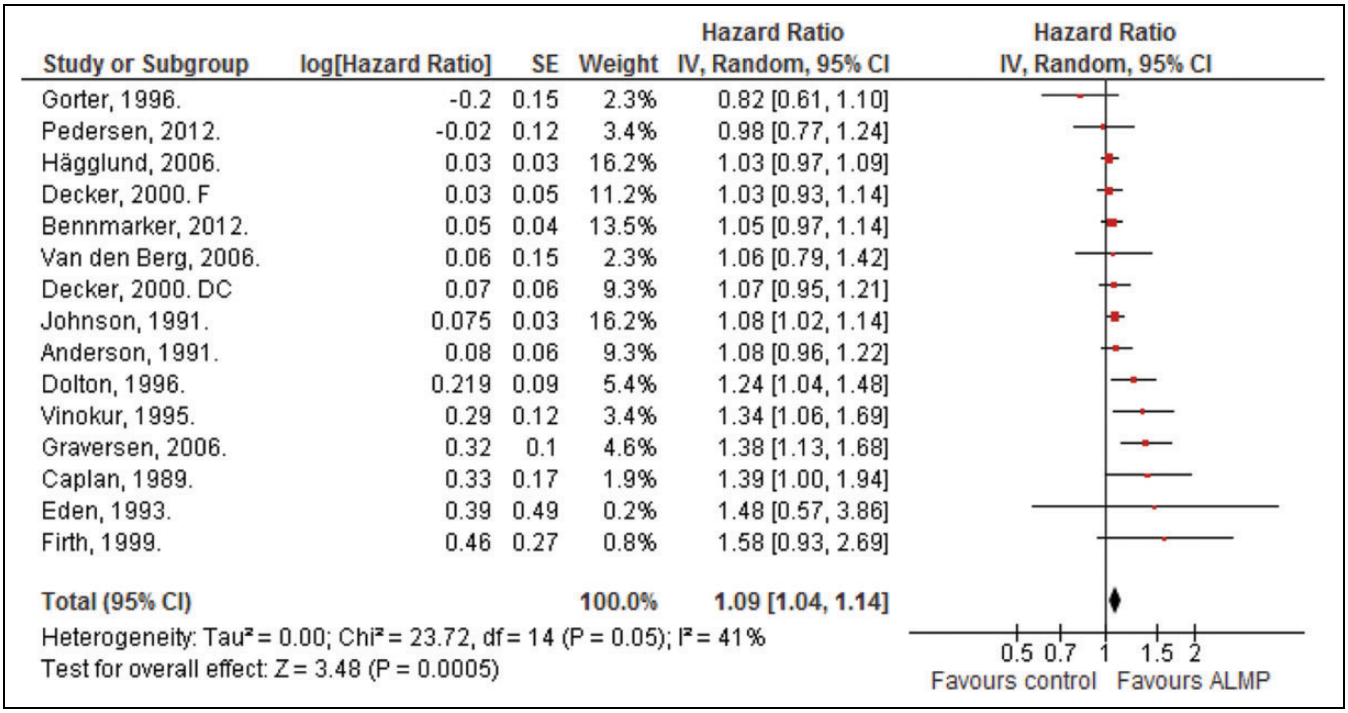

Figure 1. Forest plot, reemployment, hazard ratio.

effects (net of lock-in) of ALMP participation on the exit rate to employment. The remaining studies reported post-program effects only. We pooled effects net of lock-in and post-program effects separately for the studies using the timing-of-events approach.

We did not conduct separate analyses distinguishing between counterfactual situations, as only two studies estimated an effect relative to a world without compulsory ALMP.

Four studies reported effect measures separately for men and women. Of these, two studies further reported separate effect measures for eight different strata. One study reported separate effect measures for two former jobs. For these five studies, the average effect size was calculated and used to avoid dependence problems.

Several studies (16) reported effect measures for more than one type of ALMP and in addition four studies reported separate effect measures for different programs that were of the same type of ALMP according to our classification (see the section Types of Interventions). For these 16 studies, the average effect size was also calculated and used to avoid dependence problems.

Primary Outcome Results

Post effect measured by HRs. Fourteen studies provided in total 15 effect estimates measured as HRs post participation. The majority of reported results indicated a positive effect; only two of the study-level effects favored the control group. Ten of the study-level effects were statistically nonsignificant; only five of the study-level effects were statistically significant. Pooled results showed a significant effect. The random effects weighted mean HR was 1.09 (95% CI [1.04, 1.14], $p = .0005$). Heterogeneity of effects among studies was ignorable ($\tau^2 = .00$, $Q = 23.72$, $df = 14$, $p = .05$). See Figure 1 for the displayed forest plot.

Post effect measured by risk difference. Fourteen studies provided in total 18 effect estimates measured as risk difference post participation. The majority of reported results indicated a positive effect; only four of the study-level effects favored the control group. Thirteen of the study-level effects were statistically significant; only four of the study-level effects were statistically nonsignificant. Pooled results showed a significant effect. The random effects weighted mean risk difference was 0.07 (95% CI [0.03, 0.11], $p = .0001$). Heterogeneity of effects among studies was ignorable, although it was statistically significant ($\tau^2 = .01$, $Q = 274.87$, $df = 17$, $p < .00001$). See Figure 2 for a forest plot of reemployment and risk difference.

Net of lock in effect using the timing-of-event approach. Eight studies using the timing-of-event approach provided effect estimates net of lock in effects. The evidence is mixed; half of the reported results (4) indicated a positive effect and half of the reported results indicated a negative effect. Six of the study-level effects were statistically significant and two of the study-level effects were statistically nonsignificant. Pooled results showed a nonsignificant effect. The random effects weighted mean HR was 0.87 (95% CI [0.61, 1.25], $p = .46$. There were statistically significant heterogeneity of effects among studies ($\tau^2 = .26$, $Q = 373.03$, $df = 7$, $p < .00001$). The forest plot is displayed in Figure 3.

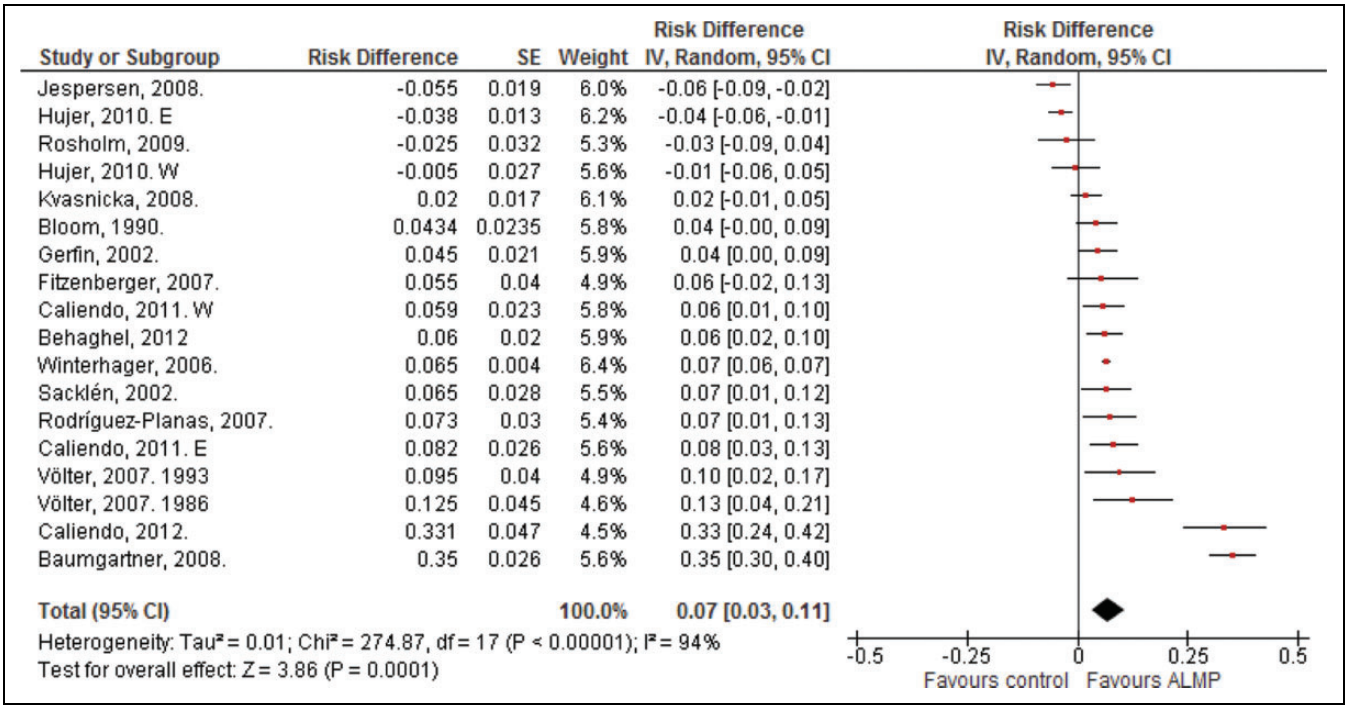

Figure 2. Forest plot, reemployment, risk difference.

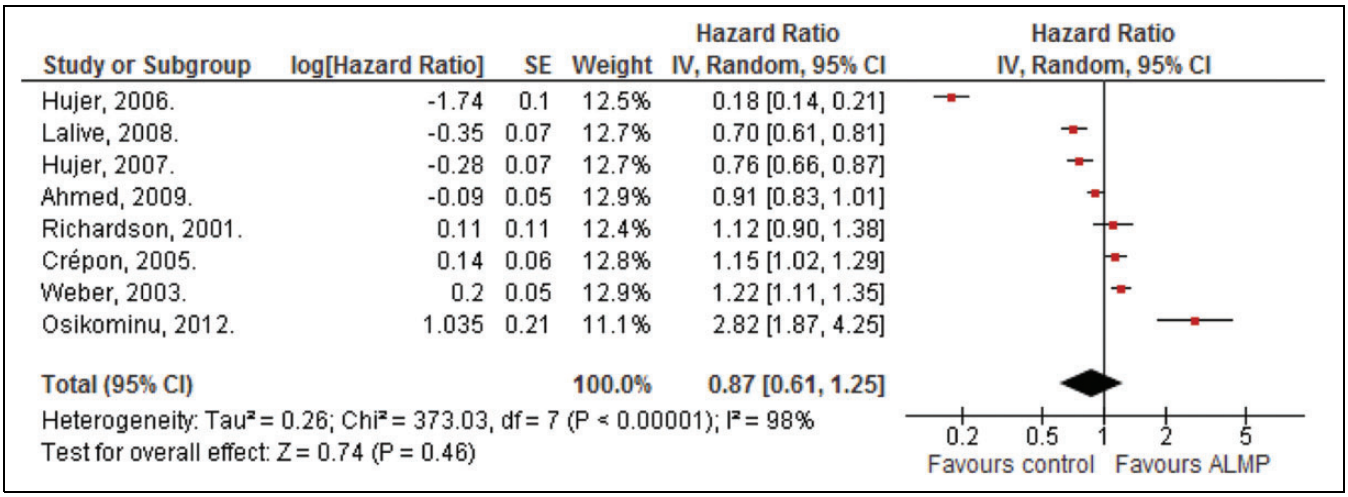

Figure 3. Forest plot, reemployment, timing-of-event, net of lock-in.

Post effect using the timing-of-event approach. Nine studies using the timing-of-event approach provided effect estimates post participation. The evidence is mixed; five of the reported results indicated a positive effect and four of the reported results indicated a negative effect. Seven of the study-level effects were statistically significant and two of the study-level effects were statistically nonsignificant. Pooled results showed a nonsignificant effect. The random effects weighted mean HR was 1.15 (95% CI [0.88, 1.49], $p = .30$). There were statistically significant heterogeneity of effects among studies ($\tau^2 = .15$, $Q = 450.87$, $df = 8$, $p < .00001$). The forest plot is displayed in Figure 4.

Summary of Primary Outcome Results

The primary outcome, the employment impact of ALMP, was analyzed separately as the effect on the exit rate to work of being assigned to ALMP at a particular moment, respectively, as the effect on employment of being assigned to ALMP in general.

The data synthesis for the effect of being assigned to ALMP in general revealed a small and statistically significant effect favoring ALMP participation. Using the timing-of-events approach, no significant effects were found of neither net of lock-in nor the post effect.

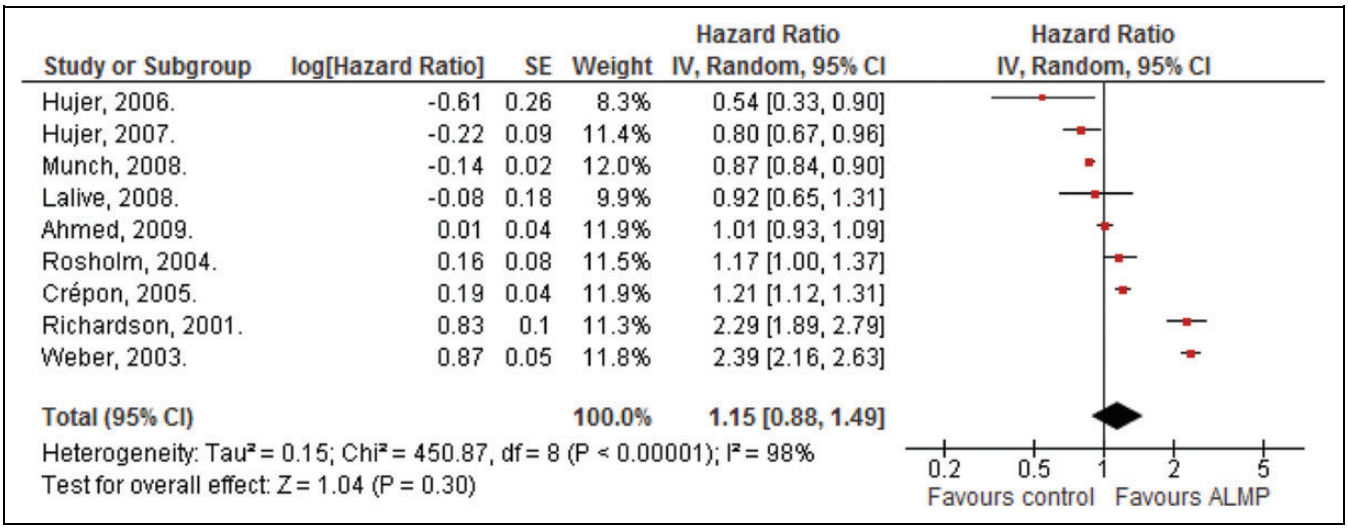

Figure 4. Forest plot, reemployment, timing-of-event, post participation.

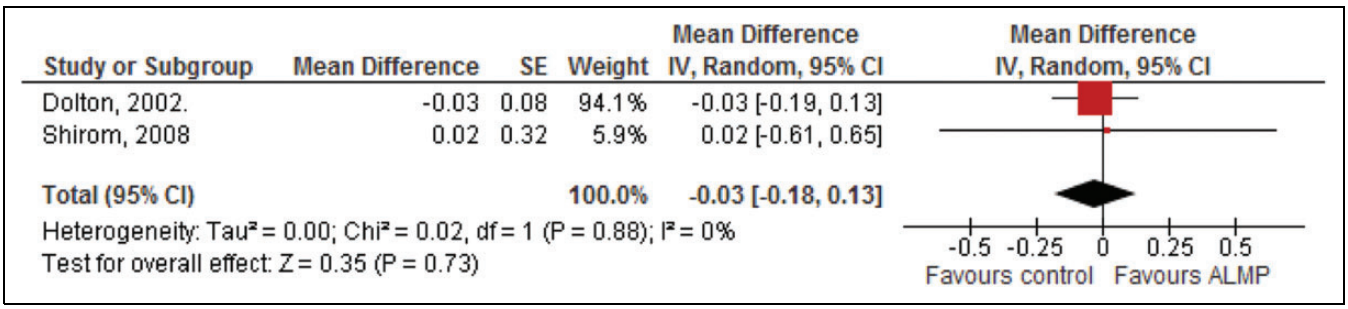

Figure 5. Forest plot, duration of reemployment.

Secondary Outcome Results

In addition to the primary outcome, we considered secondary outcomes that are relevant to the impact ALMP has on reemployment. Results on the exit rate from reemployment, duration of reemployment, and income were provided.

Two studies, Graversen (2006b) and Crépon (2005), provided data on the exit rate from reemployment. Crépon (2005) used the timing-of-events approach and so results were not pooled. Graversen (2006b) reported a nonsignificant HR of 0.98 (95% CI [0.89, 1.08], $p = .69$). Crépon (2005), using the timing-of-events approach, reported a significant HR favoring ALMP of 0.47 (95% CI [0.41, 0.54], $p < .00001$). An HR of less than 1 indicates that ALMP participation is favored. That is, the conditional exit rate from reemployment into unemployment is lower for persons who found a job after participating in ALMP than for persons who found a job without participation in ALMP.

Two studies provided data on the duration of reemployment. The evidence is inconclusive; one study reported results indicating a positive effect and one study reported results indicating a negative effect. Both of the study-level effects were statistically nonsignificant. Pooled results showed a nonsignificant effect. The random effects weighted mean difference were −0.03 (95% CI [−0.18, 0.13], $p = .73$). There were no statistically significant heterogeneity of effects among studies ($\tau^2 = .00$, $Q = .02$, $df = 18$, $p = .88$). Although the p value of the Q-statistic is notoriously underpowered to detect heterogeneity in small meta-analyses, the estimated $\tau^2 = .00$ and $I^2 = 0\%$, implying that heterogeneity among these two studies is not present. The forest plot is displayed in Figure 5.

One study, Caplan (1989), provided data on earnings that permitted the calculation of an effect size (monthly earnings for those reemployed and standard deviation). The result was a nonsignificant SMD of 0.06 (95% CI [−0.16, 0.29], $p = 0.58$).

In addition, 22 studies provided data on earnings; however, not enough information was given to calculate a SMD. The majority of the 22 studies reported an effect estimate and standard error in local currency.

Moderator Analysis and Investigation of Heterogeneity

We investigated the impact of ALMP type. Several studies (16) provided results separated by type of ALMP. We included all studies in the subgroup analyses and studies providing results for more than one type of ALMP contributed to more than one subgroup.

The risk difference post participation outcome was, in addition, investigated using meta-regression. The model was

estimated using the robust standard error method (Hedges, 2010). A random effects model in which study average effect sizes vary across studies and the effect sizes within each study are equicorrelated was used.

Subgroup Analysis

Post effect measured by HRs. It was not possible to investigate the impact of ALMP type. Only 2 of the 14 studies that provided a total of 15 effect estimates measured as HRs post participation reported results separated by ALMP type. There was no variation in the type of ALMP among the remaining studies; they were all classified as job search assistance.

Post effect measured by risk difference. Of the 15 studies providing in total 18 effect estimates measured as risk difference post participation, 6 studies reported results separated by ALMP type. Twenty-eight effect estimates were available for subgroup analysis. In addition, two effect estimates could not be classified as one of the four categories and were not included in the analysis.

The forest plot for the 28 effect estimates is displayed in Figure 6. Pooled results for the four subgroups showed a statistically significant positive effect; risk difference = 0.11 (95% CI [0.05, 0.18]) for private sector programs and nonsignificant effects; risk difference = 0.05 (95% CI [−0.02, 0.13]) for labor market training; risk difference = 0.04 (95% CI [−0.01, 0.08]) for direct employment programs in the public sector; and risk difference = 0.02 (95% CI [−0.09, 0.12]) for job search assistance. There was a statistically significant heterogeneity of effects among studies in all four subgroups ($\tau^2 = .01$, $Q = 74.06$, $df = 6$, $p < .00001$) for labor market training; ($\tau^2 = .01$, $Q = 225.00$, $df = 7$, $p < .00001$) for private sector programs; ($\tau^2 = .00$, $Q = 40.01$, $df = 7$, $p < .00001$) for direct employment programs in the public sector; and ($\tau^2 = .01$, $Q = 139.64$, $df = 4$, $p < .00001$) for job search assistance. The CIs of the subgroups overlapped.

None of the coefficients of the meta-regression were statistically significant (see Table 4). The left-out ALMP type was labor market training. An increase in effect size was seen for private sector programs, but this finding was not statistically significant (95% CI [−0.08, 0.22]). There were no significant differences in effect sizes for direct employment programs in the public sector (95% [CI −0.07, 0.07]) and for job search assistance (95% CI [−0.08, 0.06]). The estimated heterogeneity of effects among studies was small ($\tau^2 = .01$). The available evidence does not suggest that the effect of ALMP participation differs by type of ALMP.

Net of lock in effect using the timing-of-event approach. Of the eight studies using the timing-of-event approach providing effect estimates net of lock in effects, four studies reported results separated by ALMP type. Fourteen effect estimates were available for subgroup analysis. In addition, one effect estimate could not be classified as one of the four categories and was not included in the analysis.

The forest plot for the 14 effect estimates is displayed in Figure 7. There was only one effect estimate available for direct employment programs in the public sector, showing a significant negative effect. The HR was 0.78 (95% CI [0.71, 0.86]). Pooled results for the remaining three subgroups showed nonsignificant effects; HR = 0.89 (95% CI [0.56, 1.43]) for labor market training; HR = 1.07 (95% CI [0.71, 1.61]) for private sector programs; and HR = 1.09 (95% CI [0.75, 1.60]) for job search assistance. There was significant heterogeneity of effects among studies in all three subgroups ($\tau^2 = .34$, $Q = 365.47$, $df = 5$, $p < .00001$) for labor market training; ($\tau^2 = .12$, $Q = 36.31$, $df = 2$, $p < .00001$) for private sector programs; and ($\tau^2 = .15$, $Q = 210.49$, $df = 3$, $p < .00001$) for job search assistance.

The CIs for the subgroups differed only marginally with the exception of direct employment programs in the public sector, where the CI was narrow. The CIs of the other three subgroups were however inclusive of the CI of the subgroup of direct employment programs in the public sector. There was no evidence to suggest that the effect of ALMP participation net of lock in differs by type of ALMP.

Post effect using the timing-of-event approach. Of the nine studies using the timing-of-event approach providing effect estimates of post participation, five studies reported results separated by ALMP type. Twenty effect estimates were available for subgroup analysis. In addition, one effect estimate could not be classified as one of the four categories and was not included in the analysis.

The forest plot for the 20 effect estimates is displayed in Figure 8. Pooled results for the four subgroups showed a significant positive effect, HR = 1.29 (95% CI [1.04, 1.59]); for labor market training and nonsignificant effects for private sector programs, HR = 1.11 (95% CI [0.74, 1.68]); for direct employment programs in the public sector, HR = 0.94 (95% CI [0.77, 1.15]); and for job search assistance, HR = 1.06 (95% CI [0.74, 1.51]). There was significant heterogeneity of effects among studies in all four subgroups ($\tau^2 = .07$, $Q = 248.06$, $df = 6$, $p < .00001$) for labor market training; ($\tau^2 = .17$, $Q = 142.69$, $df = 3$, $p < .00001$) for private sector programs; ($\tau^2 = .03$, $Q = 20.22$, $df = 2$, $p < .00001$) for direct employment programs in the public sector; and ($\tau^2 = .19$, $Q = 805.64$, $df = 5$, $p < .00001$) for job search assistance.

The CIs of the subgroups overlapped. There is no evidence to suggest that the effect of ALMP participation differs by type of ALMP.

Sensitivity Analysis

Sensitivity analyses were planned to evaluate whether the pooled effect sizes were robust across study design and components of methodological quality. The majority of studies not using the timing-of-events approach and reporting HRs were RCTs and QRCTs. The majority of studies reporting risk difference and all studies using the timing-of-events approach were NRSs. For study design, we examined the robustness of

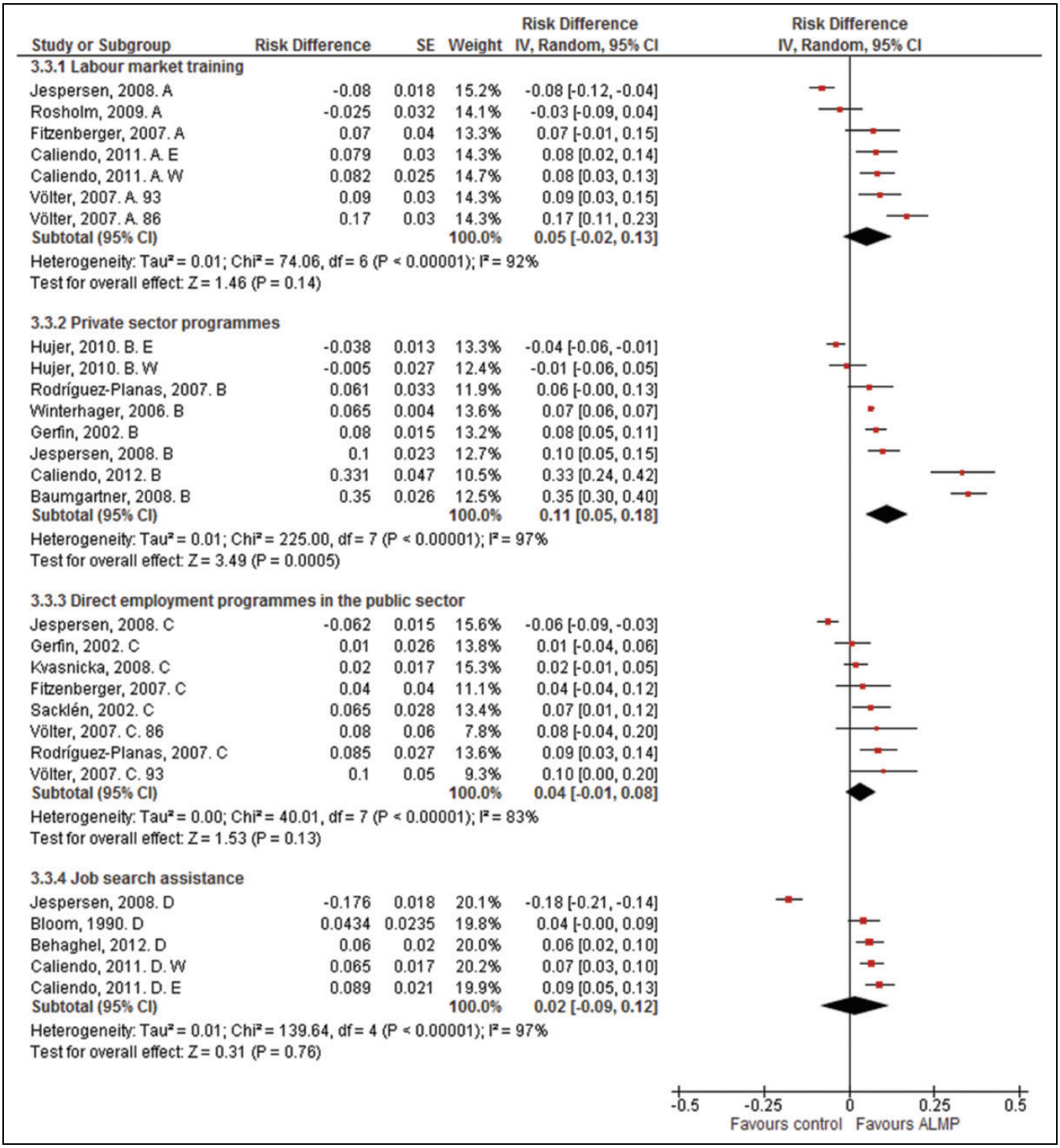

Figure 6. Forest plot, subgroups, reemployment, risk difference.

Table 4. Coefficients of Meta-Regression.

Comparison: Versus Labor Market Training	Effect Size Difference [95% CI]
Private sector programs	0.07 [−0.08, 0.22]
Direct employment programs in the public sector	−0.002 [−0.07, 0.07]
Job search assistance	−0.01 [−0.08, 0.06]

conclusions when we removed NRSs where effect sizes were measured as HRs and removal of RCTs where effect sizes were measured as risk difference. Studies using the timing-of-event were all NRSs, so we could not evaluate the impact of study design.

For methodological quality, we carried out sensitivity analyses for the allocation sequence, confounding, incomplete data, and selective reporting components of the risk of bias

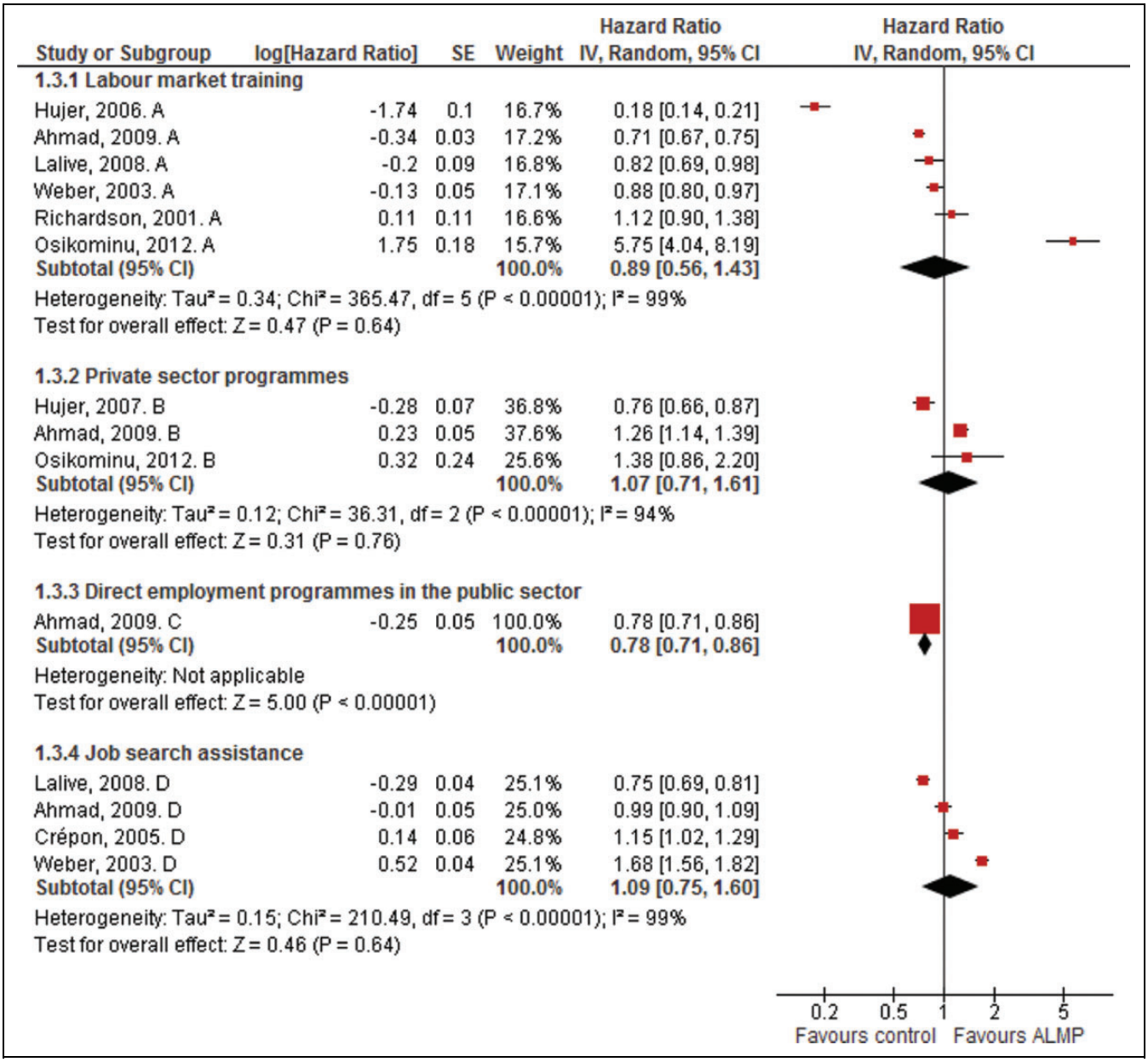

Figure 7. Forest plot, subgroups, reemployment, timing-of-event, net of lock-in.

checklists, respectively. We examined the robustness of our conclusions when we removed studies with risk of bias scores of 3 or 4 on confounding (only NRSs), incomplete data, or selective reporting. Sensitivity analyses were further used to examine the robustness of conclusions in relation to the quality of data (outcome measures based on weekly, monthly, or quarterly data collection and whether data were derived from questionnaires or administrative registers). Finally, sensitivity analyses were used to examine robustness of conclusion when we removed studies with a high (more than 25%) or unknown level of censoring. The results for studies with effects measured as HRs and risk difference are provided in Table 5 (see displayed forest plots in Filges et al., 2015).

For the studies with effects sizes measured as HRs, there was no appreciable change in the results following removal of NRSs or following removal of studies with a high/unclear risk of bias due to allocation sequence. There were no appreciable changes in the results following removal of studies with scores of 3 or 4 on the incomplete data or selective reporting components of the risk of bias checklists. Finally, there were no appreciable changes in the results following removal of studies based on quarterly data, questionnaire data, or studies with a high/unclear censoring level. The overall conclusion does not change; the hazard rate significantly increases.

For the studies with effects sizes measured as risk difference, there was no appreciable change in the results following removal of RCTs. There were no appreciable changes in the

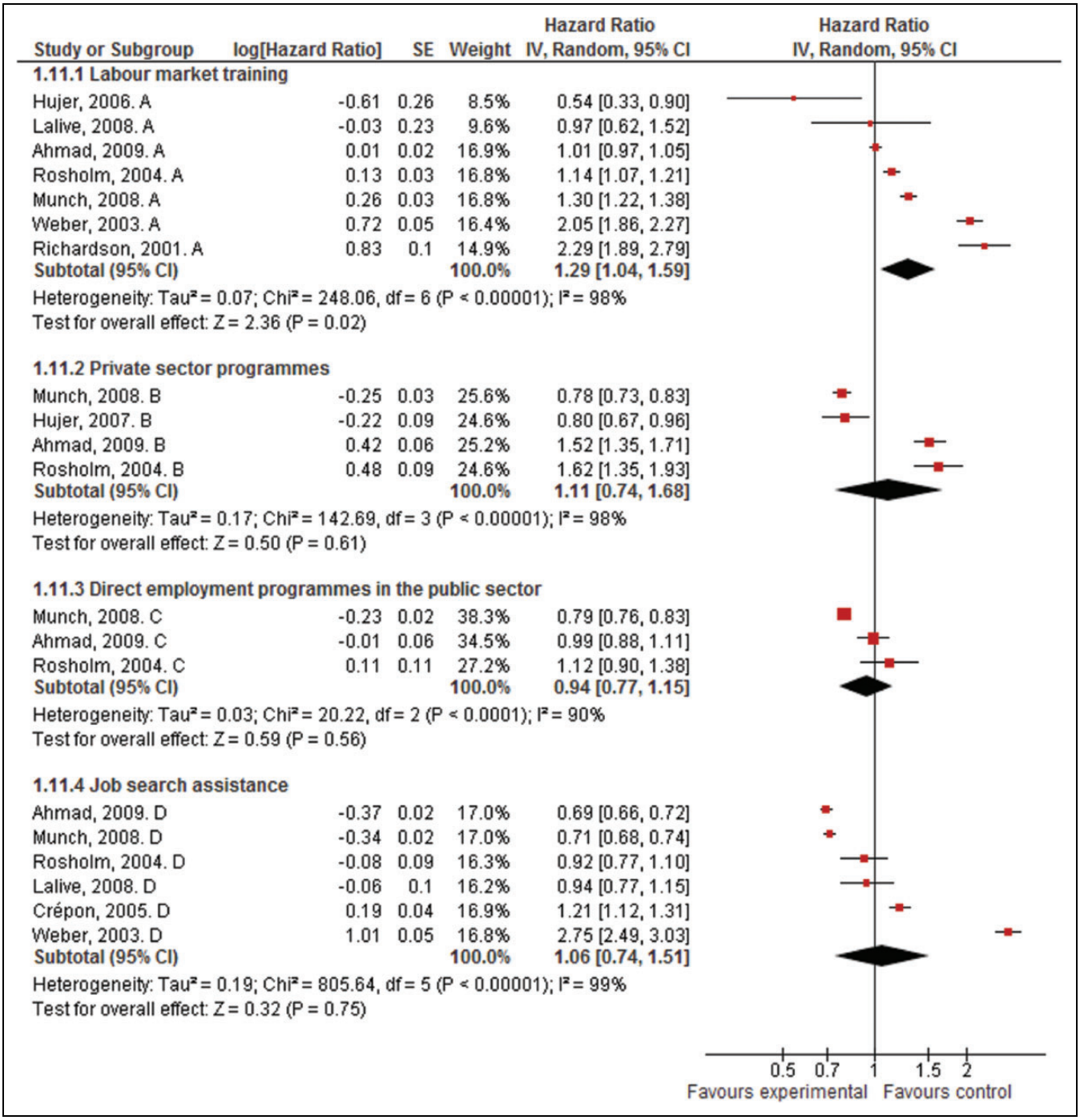

Figure 8. Forest plot, subgroups, reemployment, timing-of-event, post participation.

results following removal of studies with scores of 3 or 4 on the confounding, incomplete data, or selective reporting components of the risk of bias checklists. Finally, there were no appreciable changes in the results following removal of studies based on quarterly data or questionnaire data.

The overall conclusion does not change; the probability of employment significantly increases. The results for studies using the timing-of-event approach are provided in Table 6 (see displayed as forest plots in Filges et al., 2015).

The same pattern of results was found for the effect net of lock-in and the post effect. There were no appreciable changes in the results following removal of studies with scores of 3 or 4 on the confounding, incomplete data, or selective reporting components of the risk of bias checklists. There were no appreciable changes in the results following removal of studies with a high/unclear censoring level. The effect net of lock-in and the post effect are, however, sensitive to the removal of studies where the effect estimates were based on monthly data. The

Table 5. Sensitivity Analysis—Results for Studies With ESs Measured as HR or RD.

	Effect Size Measured as Hazard Rate	Effect Size Measured as Risk Difference
	HR [CI 95%] (Number of Studies)	RD [CI 95%] (Number of Studies)
All studies	1.09 [1.04, 1.14] (15)	0.07 [0.03, 0.11] (18)
Characteristics of studies *removed* from the analysis	ES and CI with studies removed	
RCTs	Not relevant	0.08 [0.04, 0.12] (15)
NRSs	1.09 [1.03, 1.15] (13)	Not relevant
Allocation score high/unclear	1.15 [1.03, 1.28] (4)	Not relevant
Confounding score of 4/3	Not relevant	0.07 [0.03, 0.11] (16)
Incomplete data score of 4/3	1.06 [1.01, 1.11] (7)	0.04 [0.00, 0.07] (12)
Selective reporting score of 4/3	1.10 [1.04, 1.16] (11)	0.07 [0.03, 0.11] (17)
Based on quarterly data	1.09 [1.03, 1.15] (11)[a]	0.08 [0.03, 0.12] (13)
Based on questionnaire data	1.06 [1.02, 1.10] (10)	0.04 [0.01, 0.07] (13)
High/unclear censoring level	1.09 [1.03, 1.16] (10)	Not relevant

Note. RCT = randomized controlled trial; NRS = nonrandomized study; RD = risk difference; ES = effect size; HR = hazard rate.
[a]Studies with data frequency equal to 2 months or more were excluded.

Table 6. Sensitivity Analysis—Results for Studies Using the Timing-of-Events Approach.

	Effect Net of Lock-In	Post Effect
	HR [CI 95%] (Number of Studies)	
All studies	0.87 [0.61, 1.25] (8)	1.15 [0.88, 1.49] (9)
Characteristics of studies *removed* from the analysis	HR and CI with studies removed	
Confounding score of 4/3	0.85 [0.42, 1.73] (5)	1.22 [0.67, 2.22] (5)
Incomplete data score of 4/3	0.70 [0.45, 1.08] (6)	1.13 [0.83, 1.56] (8)
Selective reporting score of 4/3	0.75 [0.52, 1.09] (7)	1.14 [0.86, 1.52] (8)
Based on monthly data	1.22 [1.00, 1.49] (5)	1.37 [1.00, 1.89] (6)
Based on questionnaire data	—	—
High/unclear censoring level	0.90 [0.56, 1.46] (2)	1.12 [0.88, 1.42] (2)

Note. "—" Indicates that no studies were based on questionnaire data. HR = hazard rate.

point estimates increase and are just significant within a 95% CI. All CIs overlap with the CIs using all studies, and so the overall conclusion remains.

Publication Bias

We assessed the possibility of publication bias visually by examining funnel plots (see the funnel plots in Filges et al., 2015). There are too few studies and insufficient variation in the standard errors to assess whether the funnel plots are symmetric. However, there is no striking asymmetry visible in any of the funnel plots.

Discussion

This review focused on the effect of participating in ALMP. The findings are mixed, depending on the approach used to investigate the effect. Two approaches were analyzed separately: the effect of being assigned to ALMP at a particular moment (the timing-of-event approach) and the effect of being assigned to ALMP in general.

The available evidence does not suggest that there is an effect of being assigned to ALMP at a particular moment. The available evidence does, however, suggest that there is an effect of participating in ALMP, although the effect is small. We found a statistically significant effect of ALMP post participation. The post effects of ALMP participation were measured by HRs and risk difference and were investigated in separate analyses. The pooled effect estimate measured as an HR is 1.09, which translates into an increase of approximately 9% in the exit rate from unemployment into employment. The pooled effect estimate measured as risk difference is an increase of 7% points in the probability of being employed approximately 1-year post participation.

In the context of HRs (the ratio of two hazard rates), the hazard is the rate within a short time interval at which the unemployed individual finds a job conditional on staying unemployed. In other words, the probability of finding a job in that short time interval is the hazard rate. The interpretation of an HR greater than 1 is that a treated unemployed person who has not yet found a job by a certain time has a higher chance of finding a job at the next point in time compared to someone in the control group.

There is an alternative interpretation of the HR that may be intuitively easier to understand. The HR is equivalent to the odds that an individual in the group with the higher hazard reaches the endpoint (finds a job) first.

Stated another way, for any pair of unemployed people, one from the treatment group and one from the control group, the HR is the odds that the time to find a job is less in the unemployed from the treatment group than in the unemployed from the control group. The probability of finding a job first (P) can easily be derived from the odds or HR of finding a job first, which is the probability of finding a job first divided by the probability of not finding a job first: HR = odds = $P/(1 - P)$

and $P = HR/(1 + HR)$ (Spotswood, Reid, Grace, & Samore, 2004). An HR of 1.09 therefore corresponds to a 52% chance of the treated unemployed person finding a job first. The lower and upper 95% CI corresponds to 51, respectively, 53% chance of the treated unemployed person finding a job first.

For interpretation of the effect measured as risk difference, we apply the number needed to treat, defined as 1/risk difference. The number needed to treat indicates, in the present context, how many unemployed people have to receive ALMP to produce one more positive event (i.e., find a job). A risk difference of 0.07 corresponds to a number needed to treat of 15 (rounded up to the next whole number). Thus, for every 15 unemployed people who participate in ALMP, an additional unemployed person will hold a job approximately 1 year after participation. The lower and upper 95% CI corresponds to a number needed to treat of 34 and 10, respectively.

It was possible to assess the impact of four types of ALMP (labor market training, private sector programs, direct employment programs in the public sector, and job search assistance). We found no evidence to suggest that the ALMP participation effect differs between these four types of ALMP.

Concerning secondary outcomes, we analyzed the effect of ALMP participation on the subsequent exit rate from reemployment, on the earnings in reemployment, and on the duration of reemployment. Only very few studies could be used in these analyses (two, one, and two, respectively). Based on the low number of studies, the evidence was inconclusive on whether participation in ALMP has an impact on the quality of the job measured as either the exit rate of reemployment, earnings in reemployment, or duration of reemployment.

Overall Completeness and Applicability of Evidence

In this review, we included 39 studies in the data synthesis. This number is relatively low compared to the large number of studies (73) meeting the inclusion criteria. If all the 73 studies had provided an effect estimate or provided data that enabled the calculation of an effect size, the final list of usable studies in the data synthesis would have been larger (avoiding overlap of data samples, 10 additional studies could have been used in the data synthesis) which again would have provided a more robust literature on which to base conclusions. The reduction was caused by three different factors. Twenty-six of the 73 studies did not report effect estimates or provide data that would allow the calculation of an effect size. Six studies were judged to have a very high risk of bias (5 on the scale) and, in accordance with the protocol, we excluded these from the data synthesis on the basis that they would be more likely to mislead than inform. Two further studies were excluded because of overlapping samples.

The 39 studies used in the data synthesis covered the United States, United Kingdom, Austria, Sweden, Denmark, Switzerland, the Netherlands, France, Romania, Israel, and (East/West) Germany (11 countries), whereas 15 countries were represented by the 73 studies.

It was not possible to examine the impact on ALMP participation of gender, age, education, or labor market conditions. It was possible to study the impact of four types of ALMP (labor market training, private sector programs, direct employment programs in the public sector, and job search assistance).

In attempt to obtain a clearer picture of the effect of ALMP participation on the quality of the job obtained, we analyzed the subsequent exit rate from reemployment, the duration of reemployment, and the reemployment earnings as secondary outcomes. Too few studies provided sufficient data for the calculation of an effect size for these outcomes and we were unable to draw a conclusion.

Quality of the Evidence

The overall quality of evidence varied from moderate to very low, depending on how the effects were measured.

The risk of bias for each of the 47 studies from which it was possible to extract an effect size was examined using a newly developed tool for assessing risk of bias incorporating NRS. We attempted to enhance the quality of the evidence in this review by excluding studies judged to be at very high risk of bias using this tool. We believe this process excluded those studies that are more likely to mislead than inform.

Concerning the overall quality, the GRADE evidence profile (see Filges et al., 2015, for details) indicates that the quality of evidence is moderate for the post effect measured by HRs, low for the post effect measured by risk difference, and very low for the effects obtained using the timing-of-event approach.

Some downgrading of evidence was undertaken for estimates of the post effect measured by HRs where the majority of studies were RCTs. This was carried out because of risk of bias in the design of the studies (see Filges et al., 2015, for details) due to limitations in the way study authors had reported the way the randomization sequence had been generated and concealed. Apart from the problems with risk of bias, the directness, consistency, precision, and publication bias were not downgraded.

No downgrading of evidence was undertaken for estimates of the post effect measured by risk difference where the majority of studies were of nonrandomized design.

Some downgrading of evidence was undertaken for estimates using the timing-of-events approach where all studies were of nonrandomized design. The reasons were (1) there was major uncertainty in the directness of the results because the effect obtained using the timing-of-events approach is the effect of being assigned to training at a particular moment, (2) there was important unexplained inconsistency (heterogeneity) in the results, and (3) CIs were very wide.

Furthermore, we performed a number of sensitivity analyses to check whether the obtained result is robust across study design, methodological quality, and data quality. The overall conclusion did not change.

To check the robustness across methodological quality, the studies with relatively high risk of bias in sequence generation, confounding, incomplete data, and selective reporting,

respectively, were removed from the analysis. To check the robustness across data quality, studies with estimates on quarterly data were removed. Monthly data were removed in the sensitivity analysis of results based on the timing-of-event approach. In addition, studies based on questionnaire data were removed.

Potential Biases in the Review Process

We believe that all the publicly available studies on the effect of ALMP participation on employment up to the censor date were identified during the review process. However, one reference was not obtained in full text and three references await translation.

We believe that there are no other potential biases in the review process as one review author (A.D.K.) and two members of the review team (S.H.F. and T.M.S.) independently coded the included studies. Any disagreements were resolved by discussion. Further, decisions about inclusion of studies and assessment of study quality were made by two review authors (A.D.K. and T.F.) independently and disagreements resolved by discussion. Numeric data extraction was made by one review author (A.D.K.) and was checked by a second review author (T.F.).

Agreements and Disagreements With Other Studies or Reviews

To the best of our knowledge, there is currently no systematic review on the effect of ALMP participation in unemployed individuals receiving unemployment insurance benefits—the focus of this review. Several papers summarize the effect of ALMP (Card et al., 2010; Heckman et al., 1999; Kluve, 2010; Kluve & Schmidt, 2002; Martin, 2000; Martin & Grubb, 2001). However, none are systematic in their search of relevant literature and none provide a synthesis of the magnitude of the effect size, although Kluve and Schmidt (2002), Kluve (2010), and Card et al. (2010) offer a meta-analysis based on vote counting and in addition investigate the contribution of covariates such as program type, participant characteristic, and country to the probability of obtaining a statistically significant positive effect. Further, Kluve (2010) and Card et al. (2010) apply ordered probit models investigating the contribution of covariates to the probability of obtaining a statistically significant positive effect, a statistically nonsignificant effect, and a statistically significant negative effect.

The focus of all these reviews is very broad, as they target unemployed individuals receiving all types of benefits. These include unemployment insurance benefits, social assistance benefits, and benefits not related to being unemployed. In addition, they include specialized types of ALMPs that target specific groups. These include specialized youth programs, vocational rehabilitation, sheltered work programs, and wage subsidies for individuals with physical, mental, or social disabilities.

Narrative surveys of ALMP experience are given by Martin (2000) and Martin and Grubb (2001) who summarize the main results of ongoing (at that time) OECD research into the effectiveness of ALMPs. Both papers draw on earlier surveys of ALMP and a few additional evaluation studies. None of the papers draw firm conclusions regarding the effect of ALMPs but merely states that the effect is not terribly encouraging.

Heckman et al. (1999) offers a descriptive summary of approximately 98 evaluation studies conducted before 1994 from the United States and Europe. The number of evaluation studies is based on counts of the number of studies in Table 22, 24, and 25 (Heckman et al., 1999). Several of the evaluation studies use the same data sample. Their search strategy is not described. No clear pattern emerges about the effectiveness of different ALMPs.

Kluve and Schmidt (2002) summarize European evaluation studies covering ALMPs conducted from 1983 to 1999, in total 53 observations. The number of studies is not reported. If a study evaluated more than one program, treatment effect estimates for all different programs were used, and if different studies reported essentially identical evaluations (same program, same time, and same result), only one of them was used. Their search strategy is not described. Thirty-three different effect estimates of programs from Europe used in Heckman et al. (1999) are included along with an additional 20 effect estimates of European programs until 1999. The authors conclude that "In summary, the estimates from recent evaluation studies suggest that treatment effects of European ALMP are rather modest (. . .)" (Kluve & Schmidt, 2002, p. 441). This is in line with the conclusion of our review. Further, they conclude that different programs are differently effective for different individuals. We were not able to investigate this aspect, as our method of analysis is very different from the one applied in Kluve and Schmidt (2002), implying that too few studies were available for these kinds of moderator analyses.

Kluve (2010) focuses on European evaluation studies covering ALMPs that were implemented in the 1990s and the 2000s, in total 137 observations originating from 96 different evaluation studies. An evaluation study may yield more than one data point, if, for example, the study evaluates more than one type of ALMP. The search strategy is not described. They do not conclude on the overall magnitude of effect size. They only conclude on the relative likelihood of different programs to estimate a significant positive and a significant negative employment outcome and find, contrary to our findings, that the programs differ in this respect. This difference in conclusions is most likely due to the very different approaches used in our review and by Kluve (2010).

Card et al. (2010) include in their analysis program evaluations conducted between 1995 and 2007. To obtain what the authors term "a comprehensive sample of recent ALMP evaluations," they e-mailed Institute for the Study of Labor research fellows who had indicated an interest in the program area "Evaluation of labour market programmes" (in total 231) and associates of the NBER Labour Studies programme (in

total 113). For details concerning the search strategy, see Card et al. (2010, pp. F454–F455).

A total of 156 studies were received and 199 effect estimates (estimates for a specific program and participant group) from 97 studies (of which 37 were also included in Kluve, 2010) were included in the analysis. They eliminated (among other things) studies which had substantial overlap with other studies included in the sample (e.g., earlier versions of the same study). For details concerning inclusion and exclusion criteria, see Card et al. (2010, pp. F455–F456). The overall conclusion of their analysis is (in line with Kluve, 2010) that the relative likelihood of different programs to estimate a significant positive and a significant negative employment outcome differ.

The available evidence analyzed in our review does suggest that there is an effect of participating in ALMP, although the size of the effect is small. This conclusion is not in disagreement with the conclusions in the above mentioned reviews; to the extent they conclude on the overall effect, they conclude that the effect is modest.

The most recent of the reviews (Card et al., 2010; Kluve, 2010; Kluve & Schmidt, 2002) analyze the relative effectiveness of ALMP types. An overall conclusion from these three reviews is that job search assistance is relatively better, and direct employment programs in the public sector relatively worse, than other programs in terms of the likelihood of these different programs to estimate a significant positive and a significant negative employment outcome. The available evidence analyzed in our review does not suggest that there is a differential effect of different types of ALMPs. However, it should be kept in mind that the apparently different conclusions concerning relative effectiveness of type of ALMP are obtained based on very different inclusion criteria concerning participants and substantially different approaches and statistical methods.

Authors' Conclusion

Implications for Practice

In this review, we have found evidence that participation in ALMP increases the probability of finding a job. The findings are, however, mixed depending on the approach used to investigate the effect. Two approaches were analyzed separately; the effect of being assigned to ALMP at a particular moment (the timing-of-event approach) and the effect of being assigned to ALMP in general. The available evidence does not suggest that there is an effect of being assigned to ALMP participation at a particular moment, neither net of lock-in nor post participation.

The available evidence does, however, suggest that there is a post effect of participating in ALMP in general, although the impact is small. The post effects of ALMP participation were measured by HRs and risk difference in separate analyses. The overall impact of ALMP participation obtained using HRs corresponds to a 52% chance of the treated unemployed person finding a job first. The overall impact of ALMP participation obtained using risk difference corresponds to a number needed to treat of 15; that is, for every 15 unemployed people who participate in ALMP, an additional unemployed person will hold a job approximately 1 year after participation.

Overall, participation in ALMPs displays a limited potential to alter the employment prospects of the individuals they intend to help.

In addition to the primary outcome, we considered secondary outcomes that are relevant to the impact of ALMP on the duration of employment and on income.

Based on the low number of studies providing data on duration of employment and income, we found no evidence to suggest that the ALMP participation has an impact on the quality of the job in terms of duration of employment and income. Thus, we have been unable to fully investigate whether the unemployed workers who are affected may actually be worse off, in the sense that they accept "bad" jobs or they are better of being offered "good" jobs. It is an important shortcoming of the current evidence that such potential detrimental side effects have not yet been fully investigated.

It was not possible to examine a number of factors which we have reason to expect have an impact on the magnitude of the effect. Knowledge of whether the effect depends on labor market conditions or whether different programs are differently effective for different individuals may be crucial to policy makers. The results of this review, however, merely suggest that across a number of different programs there is an overall small effect of ALMP participation on job finding rates and no evidence of differential effects for different programs.

The results of this review cannot be used to give advice as to whether it is appropriate to rely on ALMPs to reduce unemployment. Three reasons can be mentioned.

First, econometric estimates of individual treatment effects merely provide partial information about the impact of participation in ALMP. Any deadweight loss and substitution effects as well as any productivity and competition effects are not considered. The deadweight loss is defined as the hirings from the target group that would have occurred also in the absence of the program. The substitution effect is defined as the extent to which jobs created for a certain category of workers simply replace jobs for other categories, because relative wage costs are changed. Reliable empirical evidence that considers all direct and indirect effects on program participants and on workers not targeted by the intervention is very difficult to generate.

Second, there was insufficient evidence to take the important aspect of lock-in effects into consideration. The lock-in effect refers to the period of participation in a program. 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 ongoing skill-enhancing activity; hence the lock-in effect. The post-program effect refers to the period after participation in a program. If the ALMP has increased the individual's employability, a rise in the job-finding rate is expected. The combination of these two effects consequently determines the net effects of ALMP participation and it may be negative if there are substantive lock-in effects even if post-program effects are positive.

Third, the threat effect of compulsory participation in ALMPs should be taken into account when deciding whether or not ALMPs can be relied on to reduce unemployment. The compulsory aspect may provide an incentive for unemployed individuals to look for and return to work prior to program participation which is sometimes referred to as the threat effect. Taking into account, the threat effect may alter the evaluation of the total effects of a given program, and this may be of potentially great importance when the cost effectiveness of such programs is evaluated (or even in cost–benefit analyses of the programs).

However, some lessons can be learned from the results of the review. It was possible to assess the impact of four types of ALMP and we found no evidence to suggest that the participation effect differs by type of ALMP. This may be of potentially great importance when the cost-effectiveness of ALMPs is evaluated. As some of the programs (e.g., subsidized work, training, and education) demand full-time participation over a long time period (e.g., several months), while other programs (e.g., job search assistance and education) are part-time and have a short duration (e.g., few days/weeks); they are not equally expensive and the least expensive programs may advantageously be selected. If the least expensive programs coincide with the programs with shortest duration, the risk of substantive lock-in effects turning the net effect of program participation negative is also minimized.

Implications for Research

In this review, we found evidence that participation in ALMP increases the probability of finding a job, although the impact is small and lock-in effects are not considered. The quality of the jobs obtained, in terms of duration and income, could not be fully investigated due to limitations in the data reporting.

The planned examination of many of the potential moderators of the ALMP participation effect was not possible, as the covariates were often not reported in the included studies. Further, many of the available studies did not provide data that permitted the calculation of an effect size. If effect sizes of these studies had been available, additional valuable information about the heterogeneous effects of ALMP participation may have resulted.

These considerations point to the need for future studies that consider lock-in effects as well as heterogeneous effects of ALMP participation. Future studies should not merely report on the statistical significance of their findings but should provide their results in sufficient detail to allow their inclusion in systematic reviews examining the magnitude of effects.

Authors' Note

The review authors are responsible for any remaining errors.

Acknowledgments

We would like to thank Dr. B. C. Reeves from the Cochrane Non-Randomised Studies Methods Group for materials and training regarding the assessment of risk of bias, Lars Pico Geerdsen and Bo Honoré for valuable discussions and comments; the Campbell methods peer-referees; and external content and methods peer-referees, for valuable and insightful comments on methods and content, during the stage of writing the review report. Last but not least, thanks to the review team for their huge work morale and excellent collaboration throughout the entire review process.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Included studies

Studies with effect estimate. This first part consists of references linked to the 47 studies which provided data that permitted the calculation of an effect size. References denoted with * is the primary reference.

Agell, S. A. (1995). Swedish labor market programs: Efficiency and timing. *Swedish Economic Policy Review*, *2*, 65–98.

Ahmad, N., & Svarer, M. (2009). The effect of sanctions and active labour market programmes on the exit rate from unemployment. School of Economics and Management, University of Aarhus, Aarhus, Denmark.

*Anderson, P., Corson, W., & Decker, P. (1991). *The New Jersey unemployment insurance reemployment demonstration project. Follow-Up report* (Unemployment Insurance Occasional Discussion paper, 91 -1). Washington, DC: US Department of Labor.

Baumgartner, H. J., & Caliendo, M. (2008). Turning unemployment into self-employment: Effectiveness of two start-up programmes. *Oxford Bulletin of Economics and Statistics*, *70*, 347–373.

Behaghel, L., Crépon, B., & Gurgand, M. (2012). *Private and public provision of counseling to job-seekers: Evidence from a large controlled experiment institute for the study of labor* (IZA Discussion Paper, 6518). Bonn, Germany: Institute for the Study of Labor.

Bennmarker, H., Nordström Skans, O., & Vikman, U. (2012a). *Workfare for the old and long-term unemployed* (Working Paper, 2012:8). Uppsala, Sweden: Uppsala University, Department of Economics.

*Bennmarker, H., Nordström Skans, O., & Vikman, U. (2012b). *Workfare for the old and long-term unemployed* (Working Paper, 2012:7). Uppsala, Sweden: IFAU Institute for Evaluation of Labour Market and Education Policy.

Black, D. A., Smith, J. A., Berger, M. C., & Noel, B. J. (2002). *Is the threat of reemployment services more effective than the services themselves? Evidence from random assignment in the Ul System* (NBER Working Paper 8825). Cambridge, MA: NBER.

*Black, D. A., Smith, J. A., Berger, M. C., & Noel, B. J. (2003). Is the threat of reemployment services more effective than the services themselves? Evidence from random assignment in the Ul System. *American Economic Review*, *93*, 1313–1327.

Blasco, S., & Rosholm, M. (2011). *The impact of active labour market policy on post-unemployment outcomes: Evidence from a social experiment in Denmark* (IZA Discussion, Paper, 5631). Bonn, Germnay: Institute for the Study of Labor.

Bloom, H. S. (1990). *Back to work: Testing reemployment services for displaced workers*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Caliendo, M., Hujer, R., & Thomsen, S. L. (2005a). *The employment effects of job creation schemes in Germany: A microeconometric evaluation* (IZA Discussion Paper, 1512). Bonn, Germany: Institute for the Study of Labor.

Caliendo, M., Hujer, R., & Thomsen, S. L. (2005b). The employment effects of job creation schemes in Germany: A microeconometric evaluation. *International Journal of Social Welfare*, *20*, 144–155.

Caliendo, M., Hujer, R., & Thomsen, S. L. (2006). Sectoral heterogeneity in the employment effects of job creation schemes in Germany. *Jahrbücher für Nationalökonomie und Statistik*, *226*, 139–179.

Caliendo, M., & Künn, S. (2012). *Getting back into the labor market: The effects of start-up subsidies for unemployed females* (IZA discussion paper, 6830). Bonn, Germany: Institute for the Study of Labor.

Caliendo, M., Künn, S., & Schmidl, R. (2011). *Fighting youth unemployment: The effects of active labor market policies* (*IZA Discussion paper, 6222*). Bonn, Germany: Institute for the Study of Labor.

*Caplan, R. D., Vinokur, A. D., Price, R. H., & Van Ryn, M. (1989). Job seeking, reemployment, and mental health: A randomized field experiment in coping with job loss. *Journal of Applied Psychology*, *74*, 759–769.

*Cockx, B. (2003). *Vocational training of unemployed workers in Belgium* (IZA Discussion paper, 682). Bonn, Germany: Institute for the Study of Labor.

Cockx, B., & Bardoulat, I. (2000). *Vocational training: Does it speed up the transition rate out of unemployment?* (Tinbergen Institute Discussion paper, 16). Amsterdam and Rotterdam: Tinbergen Institute.

Corson, W., Decker, P. T., Dunstan, S. M., Gordon, A. R., Anderson, P., & Homrighausen, J. (1989). *The New Jersey unemployment insurance reemployment demonstration project: Final evaluation report* (Unemployment Insurance Occasional Discussion paper, 89-3). Washington, DC: US Department of Labor.

Corson, W., & Haimson, J. (1996). *The New Jersey unemployment insurance reemployment demonstration project: Six-year follow-up and summary report. Revised edition* (Unemployment Insurance Occasional Paper, 96-2). Washington, DC: US Department of Labor.

Crépon, B., Dejemeppe, M., & Gurgand, M. (2005). *Counseling the unemployed: Does it lower unemployment duration and recurrence?* (Working Paper, 2005 – 27). Paris, France: Centre de Recherche en Economie et Statistique.

Decker, P. T., Olsen, R. B., Freeman, L., Klepinger, D. H., Gordon, W., & Decker, P. (2000). *Assisting unemployment insurance claimants: The long-term impacts of the job search assistance demonstration*. Washington, DC: US Department of Labor, Employment and Training Administration, Office of Workforce Security.

Dolton, P., & O'Neill, D. (1995). The impact of restart on reservation wages and long-term unemployment. *Oxford Bulletin of Economics & Statistics*, *57*, 451–470.

Dolton, P., & O'Neill, D. (1996a) Unemployment duration and the restart effect: Some experimental evidence. *Economic Journal*, *106*, 387–400.

*Dolton, P., & O'Neill, D. (1996b). The restart effect and the return to full-time stable employment. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *159*, 275–288.

Dolton, P., & O'Neill, D. (2002). The long-run effects of unemployment monitoring and work-search programs: Some experimental evidence from the U.K department of economics. *Journal of Labor Economics*, *20*, 381–403. Finance and Accounting, National University of Ireland - Maynooth.

Eden, D., & Aviram, A. (1993). Self-efficacy training to speed reemployment: Helping people to help themselves. *Journal of Applied Psychology*, *78*, 352–360.

*Firth, D., Payne, C., & Payne, J. (1999). Efficacy of programmes for the unemployed: Discrete time modeling of duration data from a matched-comparison study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *162*, 111–120.

Fitzenberger, B., & Speckesser, S. (2005a). *Employment effects of the provision of specific professional skills and techniques in Germany* (ZEW Discussion Papers, 05-77). Mannheim, Germany: ZEW - Zentrum für Europäische Wirtschaftsforschung. Center for European Economic Research.

Fitzenberger, B., & Speckesser, S. (2005b). *Employment effects of the provision of specific professional skills and techniques in Germany (IZA Discussion Paper*, 1868). Bonn, Germany: Institute for the Study of Labor.

Fitzenberger, B., & Speckesser, S. (2007). Employment effects of the provision of specific professional skills and techniques in Germany. *Empirical Economics*, *32*, 529–573.

*Fitzenberger, B., & Völter, R. (2007a). *Long-run effects of training programs for the unemployed in East Germany* (ZEW Discussion Papers, 07-009). Mannheim, Germany: ZEW-Zentrum für Europäische Wirtschaftsforschung. Center for European Economic Research.

Fitzenberger, B., & Völter, R. (2007b). *Long-run effects of training programs for the unemployed in East Germany* (IZA Discussion Paper, 2630). Bonn, Germany: Institute for the Study of Labor.

Fitzenberger, B., Osikominu, A., & Völter, R. (2006). *Get training or wait? Long-run employment effects of training programs for the unemployed in West Germany* (IZA Discussion Paper, 2121). Bonn, Germany: Institute for the Study of Labor.

Fröhlich, M., & Lechner, M. (2006). *Exploiting regional treatment intensity for the evaluation of labour market policies* (IZA Discussion Paper, 2144). Bonn, Germany: Institute for the Study of Labor.

Fröhlich, M., & Lechner, M. (2010a). Exploiting regional treatment intensity for the evaluation of labor market policies. *Journal of the American Statistical Association*, *105*, 1014–1029.

*Fröhlich, M., & Lechner, M. (2010b). *Combining matching and nonparametric IV estimation: Theory and an application to the evaluation of active labour market policies* (Discussion Paper, 2010-

21). St. Gallen, Switzerland: Department of Economics, University of St. Gallen.

Frölich, M., & Lechner, M. (2004). *Regional treatment intensity as an instrument for the evaluation of labour market policies* (IZA Discussion Paper, 1095). Bonn, Germany: Institute for the Study of Labor.

*Gerfin, M., Lechner, M., & Steiger, H. (2002). *Does subsidised temporary employment get the unemployed back to work? An econometric analysis of two different schemes* (Discussion Paper, 2002-22). St. Gallen, Switzerland: Department of Economics, University of St. Gallen.

Gerfin, M., Lechner, M., & Steiger, H. (2005). Does subsidised temporary employment get the unemployed back to work? An econometric analysis of two different schemes. *Labour Economics*, *12*, 807–835.

Gorter, C., & Kalb, G. R. J. (1996). Estimating the effect of counseling and monitoring the unemployed using a job search model. *The Journal of Human Resources*, *31*, 590–610.

Graversen, B. K., & Van Ours, J. C. (2006a). *How to help unemployed find jobs quickly: Experimental evidence from a mandatory activation program Tilburg University* (Discussion Paper, 2006–126). Washington, DC: Center for Economic Research.

*Graversen, B. K., & Van ours, J. C. (2006b). *How to help unemployed find jobs quickly: Experimental evidence from a mandatory activation program* (IZA Discussion Paper, 2504). Bonn, Germany: Institute for the Study of Labor.

Graversen, B. K., & Van Ours, J. C. (2007). *How to help unemployed find jobs quickly; Experimental evidence from a mandatory activation program* (C.E.P.R. Discussion Papers). London: Centre for Economic Policy Research.

Graversen, B. K., & Van Ours, J. C. (2008a). Activating unemployed workers works; Experimental evidence from Denmark. *Economics Letters*, *100*, 308–310.

Graversen, B. K., & Van Ours, J. C. (2008b). How to help unemployed find jobs quickly: Experimental evidence from a mandatory activation program. *Journal of Public Economics*, *92*, 2020–2035.

Graversen, B. K., & Van Ours, J. C. (2009a). *How a mandatory activation program reduces unemployment durations: The effects of distance Tilburg University* (Discussion Paper, 2009–18). Washington, DC: Center for Economic Research.

Graversen, B. K., & Van Ours, J. C. (2009b). *How a mandatory activation program reduces unemployment durations: The effects of distance (IZA Discussion Paper, 4079)*. Bonn, Germany: Institute for the Study of Labor.

Graversen, B. K., & Van Ours, J. C. (2009c). *How a mandatory activation program reduces unemployment durations; the effects of distance* (IZA Discussion Paper, 4079). Bonn, Germany: Institute for the Study of Labor.

Graversen, B. K., & Van Ours, J. C. (2011). An activation program as a stick to job finding. *Labour 25*, 167–181.

Hägglund, P. (2005). *Job-search assistance using the internet—Evidence from a Swedish randomised experiment* (Working Paper, 3/2005). Stockholm, Sweden: Swedish Institute for Social Research.

*Hägglund, P. (2006). Job-search assistance using the internet: Experiences from a Swedish randomised experiment. *International Journal of Manpower*, *27*, 434–451.

Hujer, R., Thomsen, S. L., & Zeiss, C. (2004). *The effects of vocational training programmes on the duration of unemployment in Eastern Germany* (IZA Discussion Paper, 1117). Bonn, Germany: Institute for the Study of Labor.

*Hujer, R., Thomsen, S. L., & Zeiss, C. (2006). The effects of vocational training programmes on the duration of unemployment in Eastern Germany. *Allgemeines Statistisches Archiv/Journal of the German Statistical Society*, *90*, 299–321.

*Hujer, R., & Thomsen, S. L. (2010). How do the employment effects of job creation schemes differ with respect to the foregoing unemployment duration? *Labour Economics*, *17*, 38–51.

*Hujer, R., & Zeiss, C. (2007). The effects of job creation schemes on the unemployment duration in Eastern Germany. *Zeitschrift fur Arbeitsmarkt Forschung/Journal for Labour Market Research*, *40*, 383–398.

Jespersen, S. T., Munch, J. R., & Skipper, L. (2008). Costs and benefits of Danish active labour market programmes. *Labour Economics*, *15*, 859–884.

*Johnson, T. R., & Klepinger, D. H. (1991). *Evaluation of the impacts of the Washington alternative work search experiment: Final report*. Washington, DC: US Department of Labor, Employment and Training Administration, Unemployment Insurance Service.

Johnson, T. R., & Klepinger, D. H. (1994). Experimental evidence on unemployment insurance work-search policies. *Journal of Human Resources*, *29*, 665–717.

Kvasnicka, M. (2008). *Does temporary help work provide a stepping stone to regular employment*? (Working Paper, 13843). Cambridge, MA: National Bureau of Economic Research.

Lalive, R., Van Ours, J. C., & Zweimüller, J. (2000a). *The impact of active labor market programs and benefit entitlement rules on the duration of unemployment Tilburg University* (Working Paper, 2000-41). Washington, DC: Center for Economic Research.

Lalive, R., Van Ours, J. C., & Zweimüller, J. (2000b). *The impact of active labor market programs and benefit entitlement rules on the duration of unemployment* (IZA Discussion Paper, 149). Bonn, Germany: Institute for the Study of Labor.

Lalive, R., van Ours, J. C., & Zweimüller, J. (2000c). *The impact of active labor market programs on the duration of unemployment* (Working Paper, 41). Zurich, Switzerland: Institute for Empirical Research in Economics – University of Zurich.

*Lalive, R., Van Ours, J. C., & Zweimüller, J. (2008). The impact of active labour market programmes on the duration of unemployment in Switzerland. *Economic Journal*, *118*, 235–257.

Lang, K., & Leonard, J. S. (1987). Unemployment and the structure of labor markets. *American Economic Review*, *75*, 792–805.

Munch, J. R., & Skipper, L. (2008). Program participation, labor force dynamics, and accepted wage rates. *Advances in Econometrics*, *21*, 197–262.

Osikominu, A. (2012a). *Quick job entry or long-term human capital development? The dynamic effects of alternative training schemes Department of Economics* (Working Paper, 76). Zurich, Switzerland: University of Zurich.

*Osikominu, A. (2012b). *Quick job entry or long-term human capital development? The dynamic effects of alternative training schemes* (CESifo Working Paper, 3828). Munich, Germany: CESifo Group Munich.

Osikominu, A. (2013). Quick job entry or long-term human capital development? The dynamic effects of alternative training schemes. *Review of Economic Studies*, *80*, 313–342.

Payne, J., Lissenburgh, S., White, M., & Payne, C. (1996). *Employment training and employment action: An evaluation by the matched comparison method* (Research Series, 74). London, England: Department for Education and Employment.

Pedersen, J. M., Rosholm, M., & Svarer, M. (2012). *Experimental evidence on the effects of early meetings and activation* (IZA Discussion Paper, 6970). Bonn, Germany: Institute for the Study of Labor.

Prey, H. (2000). Evaluation of training programs in St. Gallen, Switzerland. *Swiss Journal of Economics and Statistics (SJES)*, *136*, 417–432.

*Richardson, K., & Berg, G. J. (2001). The effect of vocational employment training on the individual transition rate from unemployment to work. *Swedish Economic Policy Review*, *8*, 175–213.

Richardson, K., & Berg, G. J. (2006a). *Swedish labor market training and the duration of unemployment* (IZA Discussion Paper, 2314). Bonn, Germany: Institute for the Study of Labor.

Richardson, K., & Berg, G. J. (2006b). *Swedish labor market training and the duration of unemployment* (C.E.P.R. Discussion Papers, 5895). Washington, DC: Centre for Economic Policy Research.

Richardson, K., & Van den Berg, G. J. (2002). *The effect of vocational employment training on the individual transition rate from unemployment to work* (Working Paper, 2002:8). Uppsala, Sweden: IFAU - Institute for Labour Market Policy Evaluation.

Richardson, K., & van den Berg, G. J. (2008). *Duration dependence versus unobserved heterogeneity in treatment effects: Swedish labor market training and the transition rate to employment* (Working Paper, 2008:7). Uppsala, Sweden: IFAU - Institute for Labour Market Policy Evaluation.

Rodriguez-Planas, N. (2007). *What works best for getting the unemployed back to work: Employment services or small-susiness assistance programmes?* (IZA Discussion Paper, 3051). Bonn, Germany: Institute for the Study of Labor.

*Rodriguez-Planas, N., & Benus, J. (2007). *Evaluating active labor market programs in Romania* (Working Paper, 2007-31). Madrid, Spain: FEDEA.

Rodriguez-Planas, N., & Benus, J. (2010). Evaluating active labor market programs in Romania. *Empirical Economics*, *38*, 65–84.

Røed, K., & Raaum, O. r. (2003a). *The effect of programme participation on the transition rate from unemployment to employment*. Oslo, Norway: Oslo University, Department of Economics.

*Røed, K., & Raaum, O. r. (2003b). Do labour market programmes speed up the return to work? *Oxford Bulletin of Economics & Statistics*, *68*, 541–568.

Rosholm, M., & Skipper, L. (2009). Is labour market training a curse for the unemployed? Evidence from a social experiment. *Journal of Applied Econometrics*, *24*, 338–365.

Rosholm, M., & Svarer, M. (2004a). *Estimating the threat effect of active labour market programmes* (IZA Discussion Paper, 1300). Bonn, Germany: Institute for the Study of labor.

*Rosholm, M., & Svarer, M. (2004b). *Estimating the threat effect of active labour market programmes school of economics and management* (Working Paper, 2004-06). Aarhus, Denmark: University of Aarhus.

Rosholm, M., & Svarer, M. (2008). The threat effect of active labour market programmes. *Scandinavian Journal of Economics*, *110*, 385–401.

Sacklén, H. (2002). *An evaluation of the Swedish trainee replacement schemes (Working Paper, 7)*. Uppsala, Sweden: IFAU-Institute for Labour Market Policy Evaluation.

Solie, R. J. (1968). Employment effects on retraining the unemployed. *Industrial & Labor Relations Review*, *21*, 210–225.

Sørensen, K. L. (2012). *Effects of intensifying labor market programs on post-unemployment wages: Evidence from a controlled experiment* (Economic working Papers, 2012-20). Aarhus, Denmark: School of Economics and Management, University of Aarhus.

Steinberg, D., & Monforte, F. A. (1987). Estimating the effects of job search assistance and training programs on the unemployment durations of displaced workers. In K. Lang & J. Leonard (Eds.), *Unemployment and the structure of labor markets* (pp. 186–206). New York, NY: B. Blackwell.

Van den Berg, G. J., & Van der Klaauw, B. (2001a). *Counseling and monitoring of unemployed workers: theory and evidence from a controlled social experiment (Working Paper, 2001:12)*. Uppsala, Sweden: IFAU - Institute for Evaluation of Labour Market and Education Policy.

Van den Berg, G. J., & Van der Klaauw, B. (2001b). *Counselling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment (IZA Discussion Paper 374)*. Bonn, Germany: Institute for the Study of labor.

*Van den Berg, G. J., & Van der Klaauw, B. (2006). Counseling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment. *International Economic Review*, *47*, 895–936.

Vikström, J., Rosholm, M., & Svarer, M. (2011). *The relative efficiency of active labour market policies: Evidence from a social experiment and non-Parametric Methods* (IZA Discussion Paper, 5596). Bonn, Germany: Institute for the Study of Labor.

*Vinokur, A. D., Price, R. H., & Schul, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology*, *23*, 39–74.

Vinokur, A. D., Schul, Y., Vuori, J., & Price, R. H. (2000). Two years after a job loss: Long-term impact of the JOBS program on reemployment and mental health. *Journal of Occupational Health Psychology*, *5*, 32–47.

Vinokur, A. D., van, R. M., Gramlich, E. M., & Price, R. H. (1991). Long-term follow-up and benefit-cost analysis of the jobs program: A preventive intervention for the unemployed. *The Journal of Applied Psychology*, *76*, 213–219.

*Völter, R., Osikominu, A., & Fitzenberger, B. (2007). *Get training or wait? Long-run employment effects of training programs for the unemployed in West Germany* (ZEW Discussion Papers, 06-39). Mannheim, Germany: ZEW - Zentrum für Europäische

Wirtschaftsforschung / Center for European Economic Research.

Weber, A., & Hofer, H. (2004a). *Employment effects of early interventions on job search programs (IZA Discussion Paper*, 1076). Bonn, Germany: Institute for the Study of Labor.

*Weber, A., & Hofer, H. (2004b). *Are job search programs a promising tool? A microeconometric evaluation for Austria* (IZA Discussion Paper, 1075). Bonn, Germany: Institute for the Study of Labor.

Winterhager, H., Heinze, A., & Spermann, A. (2006). Deregulating job placement in Europe: A microeconometric evaluation of an innovative voucher scheme in Germany. *Labour Economics*, *13*, 505–517.

Studies without effect estimate: This second part consists of references linked to the 26 studies which did not provide data that permitted the calculation of an effect size. References indicated with * is the primary reference.

Adda, J., Costa Dias, M., Meghir, C., & Sianesi, B. (2007). *Labour market programmes and labour market outcomes: A study of the Swedish active labour market interventions IFAU* (Working Paper, 2007:27*)*. Uppsala, Sweden: Institute for Labour Market Policy Evaluation.

Beenstock, M. (1996). Training and the time to find a job in Israel. *Applied Economics*, *28*, 935–946.

Carling, K., & Richardson, K. (2001). *The relative efficiency of labor market programs: Swedish experience from the 1990s* (IZA Discussion Paper, 3767). Uppsala, Sweden: IFAU - Institute for Labour Market Policy Evaluation.

*Carling, K., & Richardson, K. (2004). The relative efficiency of labor market programs: Swedish experience from the 1990s. *Labour Economics*, *11*, 335–354.

Cavaco, S., Fougère, D., & Pouget, J. (2005). *Estimating the effect of a retraining program for displaced workers on their transition to permanent jobs* (IZA Discussion Paper, 1513). Bonn, Germany: Institute for the Study of Labor.

*Cockx, B., & Van der Linden, B. (1996). Active labour market policies and job tenure. *Oxford Economic Papers*, *50*, 685–708.

Cockx, B., Van der Linden, B., & Karaa, A. (1996). *Active labour market policies and job tenure*. Leuven, Belgium: University catholique de Louvain. Institut de Recherches Economiques et Sociales (IRES).

Forslund, A., Johansson, P., & Lindqvist, L. (2004). *Employment subsidies-a fast lane from unemployment to work?* (Working Paper, 2004:18). Uppsala, Sweden: Institute for Labour Market Policy Evaluation (IFAU).

Fredriksson, P., & Johansson, P. (2003). *Employment, mobility, and active labor market programs (Working Paper, 2003:3)*. Uppsala, Sweden: Employment, mobility, and active labor market program (IFAU).

Gerfin, M., & Lechner, M. A. (2000). *Microeconometric evaluation of the active labour market policy in Switzerland* (IZA Discussion Paper, 154). Bonn, Germany: Institute for the Study of Labor.

Gerfin, M., & Lechner, M. A. (2001). *Microeconometric evaluation of the active labour market policy in Switzerland* (C.E.P.R. Discussion Papers, 2934). Washington, DC: Centre for Economic Policy Research.

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

Hanna, J., Turney, Z., & Williams, L. S. (1990). *UI research exchange* (Unemployment Insurance Occasional Paper, 90-4). Washington, DC: U.S. Department of Labor.

Klepinger, D. H., Johnson, T. R., Joesch, J. M., & Benus, J. M. (1997). *Evaluation of the Maryland unemployment insurance work search demonstration* (UI Occasional Paper, 98-2), Baltimore, MA: Maryland Department of Labor, Licensing and Regulation, Office of Unemployment Insurance.

Lechner, M., Miquel, R., & Wunsch, C. (2004). *Long-run effects of public sector sponsored training in West Germany* (IZA Discussion Paper, 1443). Bonn, Germany: Institute for the Study of Labor.

Lechner, M., Miquel, R., & Wunsch, C. (2005). *Long-run effects of public sector sponsored training in West Germany* (C.E.P.R. Discussion Papers, 4851). Washington, DC: Centre for Economic Policy Research.

*Lechner, M., Miquel, R., & Wunsch, C. (2011). Long-run effects of public sector sponsored training in West Germany. *Journal of the European Economic Association*, *9*, 742–784.

Lechner, M., & Wiehler, S. (2007a). *Does the order and timing of active labor market programs matter?* (IZA Discussion Paper, 3092). Bonn, Germany: Institute for the Study of Labor.

Lechner, M., & Wiehler, S. (2007b). *Kids or courses? Gender differences in the effects of active labour market policies* (IZA Discussion Paper, 2740). Bonn, Germany: Institute for the Study of Labor.

Lechner, M., & Wunsch, C. (2006a). *Are training programs more effective when unemployment is high?* (IZA Discussion Paper, 2355). Bonn, Germany: Institute for the Study of Labor.

Lechner, M., & Wunsch, C. (2006b). *Active labour market policy in East Germany: Waiting for the economy to take off* (IZA Discussion Paper, 2363). Bonn, Germany: Institute for the Study of Labor.

Lechner, M., & Wunsch, C. (2007). *What did all the money do? On the general ineffectiveness of recent West German labour market programmes* (IZA Discussion Paper, 2800). Bonn, Germany: Institute for the Study of Labor.

Lechner, M., & Wunsch, C. (2009a). Are training programs more effective when unemployment is high? *Journal of Labor Economics*, *27*, 653–692.

*Lechner, M., & Wunsch, C. (2009b). Active labour market policy in East Germany. *Economics of Transition*, *17*, 661–702.

O'Leary, C. J. (1995). *An impact analysis of labor market programs in Hungary* (Upjohn Institute Working Paper, 95-30). Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

*O'Leary, C. J. (1997). A net impact analysis of active labour programmes in Hungary. *Economics of Transition*, *5*, 453–484.

O'Leary, C. J., Kolodziejczyk, P., & Lazar, G. (1998). The net impact of active labour programmes in Hungary and Poland. *International Labour Review*, *137*, 321–346.

Raaum, O. r., Torp, H., & Zhang, T. (2003). *Business cycles and the impact of labour market programmes* (Memorandum, 14/2002). Oslo, Norway: Oslo University, Department of Economics.

Shirom, A., Vinokur, A., & Price, R. (2008). Self-efficacy as a moderator of the effects of job-search workshops on re-employment: A field experiment. *Journal of Applied Social Psychology*, *38*, 1778–1804.

Sianesi, B. (2001). *Differential effects of Swedish active labour market programmes for unemployed adults during the 1990s* (IZA Discussion Paper, 5631). Bonn, Germany: Institute for the Study of Labor.

Sianesi, B. (2002). *Differential effects of Swedish active labour market programmes for unemployed adults during the 1990s* (Working Paper, 2002:5). Uppsala, Sweden: IFAU - Institute for Labour Market Policy Evaluation.

*Sianesi, B. (2008). Differential effects of active labour market programs for the unemployed. *Labour Economics*, *15*, 370–399.

Steiger, H. (2004). *Is less more? A look at nonparticipation in Swiss active labour market programmes* (Unpublished working paper). St.Gallen, Switzerland: Swiss Institute for International Economics and Applied Economic Research, University of St.Gallen.

Stephan, G., & Pahnke, A. (2008). *The relative effectiveness of selected active labour market programmes and the common support problem* (IZA Discussion Paper, 3767). Bonn, Germany: Institute for the Study of Labor.

Torp, H. (1994). The impact of training on employment - Assessing a Norwegian labor-market program. *Scandinavian Journal of Economics*, *96*, 531–550.

Vuori, J., Silvonen, J., Vinokur, A. D., & Price, R. H. (2002). The Työhön job search program in Finland: Benefits for the unemployed with risk of depression or discouragement. *Journal of Occupational Health Psychology*, *7*, 5–19.

Vuori, J., & Vesalainen, J. (1999). Labour market interventions as predictors of re-employment, job seeking activity and psychological distress among the unemployed. *Journal of Occupational and Organizational Psychology*, *72*, 523–538.

*Wunsch, C., & Lechner, M. (2008). What did all the money do? On the general ineffectiveness of recent West German labour market programmes. *Kyklos*, *61*, 134–174.

Zhang, T. (2003). *Identifying treatment effects of active labour market programmes for Norwegian adults* (Memorandum, 26/2003). Oslo, Norway: Oslo University, Department of Economics.

Additional references

Aarnio, O. (1996). Labour market policies in the Nordic countries – Some aggregate-level empirical evidence. In E. Wadensjö (Ed.), *The Nordic labour markets in the 1990's* (pp. 203–246). North-Holland, the Netherlands: Elsevier.

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., . . . Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, *64*, 401–406.

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons.

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

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

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

Filges, T., Geerdsen, L. P., Knudsen, A. D., & Jørgensen, A. K. (2014). Unemployment benefit exhaustion incentive effects on job-finding rates. *Research on Social Work Practice*, *25*, 1–23.

Filges, T., Geerdsen, L. P., Smedslund, G., Knudsen, A. D., & Jørgensen, A. K. (2013). Active labour market programme participation for unemployment insurance recipients: Protocol for a systematic review. *Campbell Systematic Reviews*. Retrieved from http://campbellcollaboration.org/lib/project/185/

Filges, T., Geerdsen, L. P., Smedslund, G., Knudsen, A. D., & Jørgensen, A. K. (2015). Active labour market programme participation for unemployment insurance recipients: A systematic review. *Campbell Systematic Reviews*, *11*, 1–344.

Filges, T., & Hansen, A. T. (in press). The threat of active labour market programmes for unemployed. Forthcoming in *Journal of Economic Surveys*.

Filges, T., Kennes, J., Larsen, B., & Tranæs, T. (2011). Labour market programmes and the equity-efficiency trade-off. *Journal of Macroeconomics*, *33*, 738–753.

Geerdsen, L. P. (2003). *Marginalisation processes in the Danish labour market* (Ph.D. thesis). Copenhagen, Denmark: The Danish National Institute of Social Research, Report 03:24.

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

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., & Alonso-Coello, P., . . . GRADE Working Group. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, *336*, 924–926.

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

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

Hedges, L. V. (2007). Meta-analysis. In C. R. Rao (Eds.), *The handbook of statistics* (pp. 919–953). Amsterdam, the Netherlands: Elsevier.

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

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0) [updated March 2011]. Hoboken, NJ: Wiley-Blackwell The Cochrane Collaboration. Retrieved from www.cochrane-handbook.org

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560.

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

Kluve, J. (2006). *The effectiveness of European active labor market policy* (IZA Discussion Paper, 2018). Retrieved from http://ftp.iza.org/dp2018.pdf

Kluve, J. (2010). The effectiveness of European active labor market programs. *Labour Economics 17*, 904–918.

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

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

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

Lancaster, T. (1990). *The econometric analysis of transition data*. Cambridge, MA: Cambridge University Press.

Layard, R., Nickell, S., & Jackman, R. (2005). *Unemployment - Macroeconomic performance and the labour market* (2nd ed.). Oxford, England: Oxford University Press.

Martin, J. P. (2000). What works among active labour market policies: Evidence from OECD countries' experiences. *OECD Economic Studies*, *30*, 79–113.

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

Mortensen, D. T. (1987). Job search and labor market analysis. In O. Ashenfelter & R. Layard (Eds.), *Handbook of labor economics* (pp. 849–919). Amsterdam, the Netherlands: Elsevier.

OECD. (2004). *Employment outlook*. Paris, France: OECD.

OECD. (2007). *Benefits and wages*. Paris, France: OECD Indicators.

Oxman, A., & Guyatt, G. H. (1992). Consumer's guide to subgroup analyses. *Annals of Internal Medicine*, *116*, 78–84.

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

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

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–50.

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

Smedslund, G., Hagen, K. B., Steiro, A., Johme, T., Dalsbø, T. K., & Rud, M. G. (2006). Work programmes for welfare recipients. *Campbell Collaboration Review*, *2*, 1–125.

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

Spotswood, L. S., Reid, J. E., Grace, M., & Samore, M. (2004). Hazard ratio in clinical trials. *Antimicrobial. Agents Chemotherapy*, *48*, 2787–2792.