JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12134

THE THREAT EFFECT OF ACTIVE LABOR MARKET PROGRAMS: A SYSTEMATIC REVIEW

Trine Filges*

The Danish National Centre for Social Research SFI Campell

Anne Toft Hansen

The Danish National Centre for Social Research School and Education

Abstract. This paper is a systematic review of the threat effect of active labor market programs for unemployed individuals. The threat effect is the induced change in the hazard rate of leaving unemployment prior to program participation. Studies included in the review all estimated a threat effect, with the participants in all cases being unemployed individuals in receipt of benefit of some kind during their tenure of unemployment. Eight of these studies have been included in a meta-analysis: The meta-analysis, which has been carried out using a random effects model to account for heterogeneity, indicated a hazard ratio of 1.25 for the pooled estimate. We conclude that active labor market programs constitute a statistically significant threat effect, although it is modest.

Keywords. Active labor market programs; Hazard ratio; Systematic review; Threat effect

1. Introduction

Active labor market programs (ALMPs) were introduced during the 1990s as national public policy in various countries with the aim of reducing unemployment. In some countries, such as Australia, the USA, Denmark, Sweden, the UK, and Switzerland, participation in an ALMP is required in order to continue receiving benefits (Gerfin and Lechner, 2002; Geerdsen, 2006). Typically, compulsory program participation is required after a specific duration of unemployment. The purpose of making benefit payments conditional on participation in ALMPs is twofold. First, it may improve the participants' qualifications and reintroduce them to the labor market. Second, the compulsory aspect may provide an incentive for unemployed individuals to look for and return to work and thereby mitigate the occurrence of moral hazard (Jackman, 1994; Hansen and Tranæs, 1999; Black *et al.*, 2003).

This paper reviews studies that focus on the second effect; we analyze the effect that occurs *prior* to participation in a compulsory ALMP. It is the threat of intervention that creates the effect, for example, the threat of the ALMP that the unemployed individuals have to participate in, in order to continue receiving benefits (we denote this effect prior to participation as the "threat effect").

The effects during and after program participation have extensively been investigated and reviewed (Heckman *et al.*, 1999; Martin and Grubb, 2001; Filges *et al.*, 2015). Often, the effects of ALMPs are

*Corresponding author contact email: tif@sfi.dk; Tel: +45 33 48 09 26.

found to be close to zero. If there is a threat effect of ALMPs, then it should be taken into account when assessing the total effect of ALMPs for policy purposes, especially if the threat effect is the only effect of the programs. The existence of the threat effect has only recently been recognized, yet some evidence does exist of these effects (Black *et al.*, 2003). However, this research has not yet been summarized, which is vital if one is to gain a clear picture of the available evidence on threat effects. This is the purpose of this review.

The type of program that unemployed individuals are required to participate in is relevant to the threat effect, as the incentive to leave unemployment is stronger when the program is considered unattractive or burdensome. Therefore, an important aspect of the interventions is that they take up some part of the unemployed individuals' time. On the one hand, participation in the programs could act as a "leisure tax," raising the cost of continued benefit receipt (for those who do not value the programs). The threat effect would then be due to the unemployed attempting to avoid participation in the program. On the other hand, some people might value the programs, as they expect to benefit from their participation, for instance, if it consists of vocational training from which they receive a certificate or other credentials. For further discussion of the theoretical aspects, see Moffitt (1996).

We find a positive threat effect of ALMPs, highlighting the importance of taking such an effect into account when evaluating the overall effect of an ALMP.

The paper is organized as follows: Section 2 explains the selection criteria of the data collection and the methods used; Section 3 presents the final study population; Section 4 shows the results of the review and our meta-analysis; and Section 5 discusses the results and concludes. We follow the meta-analysis reporting guidelines provided by Stanley *et al.* (2013) throughout this review.

2. Selection Criteria, Results of the Search, and Method

The selection criteria for which studies to include are based on types of studies, types of participants, and types of outcomes. This section presents a summary of each selection criterion, the results of the search, and briefly describes the method of analysis.

2.1 *Selection Criteria*

2.1.1 *Types of Studies*

We have reviewed all available studies that estimate a threat effect, either using a control group or using an estimated counterfactual outcome. ALMPs are often implemented as general national policies in the countries considered. In most of these countries, no experiments have been performed regarding the efficiency of such social and labor market policies. Therefore, studies on the effects of ALMPs often have to rely on nonrandom assignment using observational data. In these cases, the effects are estimated by econometric methods using various identifying assumptions to estimate counterfactual outcomes.

2.1.2 *Types of Participants*

In order for a study to be included, the participants all had to be unemployed individuals who received some sort of benefit during their tenure of unemployment. These benefits could be unemployment insurance (UI) benefits or social assistance benefits. UI systems are organized differently in different countries, often with different rules for membership, eligibility, etc. In some countries, two systems provide benefits to unemployed individuals: A UI system for individuals who have a strong labor market attachment (i.e., who are often employed) and a social welfare system for individuals who typically have other problems in addition to unemployment. We include studies with participants in all types of unemployment benefit systems.

2.1.3 *Types of Outcomes*

As we are focused on whether the prospect of program participation motivates unemployed individuals to leave the unemployment benefit system, the primary outcome is the individuals' hazard rate of leaving unemployment. The threat effect is the difference in hazard rates prior to program participation between individuals who are required to participate and individuals who are not required to participate (or the hazard rate they would have had if they had not been required to participate).

2.2 *Results of the Search*

The literature search was conducted in October 2014. The general search strategy setup in collaboration with the SFI Campbell librarian generated 5602 hits in the different databases (see Appendix A for search methods). Two research assistants screened all the titles and abstracts. The first-level screening resulted in 64 potentially relevant titles/abstracts. Full texts were retrieved for all the potentially relevant titles/abstracts and were second level screened by two reviewers. Based upon the inclusion criteria, the selection resulted in 21 studies. Only the latest version of each study, typically the journal article version, was included.

In addition to the electronic literature search, we looked through the references of different studies that seemed relevant as well as chapters of the *Handbook of Labor Economics*: this resulted in four additional hits. Furthermore, searches were conducted for citations of the two studies by Black *et al.* (2003) and Richardson (2002). None of these searches gave any relevant hits. One additional study was located after being suggested by a colleague.

Upon inspection of the potentially relevant studies, it became apparent that six studies did not qualify for inclusion in the review (reasons for exclusions available on request). This resulted in a final selection of 20 studies from eight different countries.

2.3 *Method of Analysis*

We provide a quantitative synthesis of the results as vote counting (counting the number of significant predictors to arrive at a conclusion) has poor statistical properties because it ignores the magnitudes of the effect sizes (Hedges and Olkin, 1980). Constructing a quantitative meta-analysis requires comparable effect sizes. With few exceptions, the studies report hazard ratios. The hazard ratio measures the proportional change in hazard rates between unemployed persons approaching ALMPs and unemployed persons not approaching ALMPs.

The estimation method also influences the way the threat effect is reported. The approach shared by the majority of the studies is using indicator variables for time until compulsory participation in a program. We perform a meta-analysis based on effect estimates from the studies that use this approach using the inverse-variance-weighted random-effect model that incorporates both the sampling variances and between study variance components into the study-level weights (Hedges, 2007). See Appendices B and C for mathematical derivations.

3. Data

Table 1 depicts the studies of the final study population, consisting of 20 studies covering a range of countries: Three studies analyze ALMPs in the USA, one in Australia, one in Switzerland, one in

Table 1. Characteristics of Studies Included in the Review.

Country	Author/Year	Program type	Duration until program start	Observation period	Identification strategy	Program duration	Threat effect
United States	Benus and Johnson (1997)	UI	3–5 weeks	1994–1995	Random assignment	3–5 weeks	+
	Scrivener *et al.* (2002)	Welfare recipients	30 months	1994–2001	Random assignment	10 months	+
	Black *et al.* (2003)	UI, social benefit, etc.	3–4 weeks	1994–1996	Random assignment	Variable	+
Denmark	Jensen *et al.* (2003)	UI	6 months	1996	Exclusion restrictions	18 months	NS
	Rosholm (2008)[A]	UI	5–6 weeks	2005–2006	Random assignment	3 months	+
	Graversen and van Ours (2008)[A]	UI	5–6 weeks	2005–2008	Random assignment	3 months	+
	Pedersen *et al.* (2012)	UI	0–13 weeks	2008–2010	Random assignment	12 weeks	NS
	Geerdsen and Holm (2007)[B]	UI	4 years, 3 years, 2 years, differences between groups	1995–1998	Legislative shortening in timing of compulsory labor market programs, affecting all UI individuals	Not mentioned	+
	Rosholm and Svarer (2008)	UI	Varies around 52 weeks	1998–2002	Functional form assumptions on hazard rate out of unemployment	Variable. A few weeks to more than a year	+
	Kyhl (2001)[B]	UI	2–3 years	1996–1998	Legislative shortening in timing of compulsory labor market programs, affecting all UI individuals	36 months	+

(Continued)

Table 1. *Continued*

Country	Author/Year	Program type	Duration until program start	Observation period	Identification strategy	Program duration	Threat effect
	Geerdsen (2006)[B]	UI	4 years, 3 years, 2 years, differences between groups	1994–1998	Legislative shortening in timing of compulsory labor market programs, affecting all UI individuals	75–80% of the time	+
	Clausen *et al.* (2013)[B]	UI	12–21 months	1998–2001	Legislative shortening in timing of compulsory labor market programs, affecting all UI individuals	3 years	+
	Amilon (2010)[B]	UI	1 year	1991–1999	Legislative shortening in timing of compulsory labor market programs, affecting all UI individuals	3 years	+
	Graversen (2004)	Social benefits	0–52 weeks	1994–1998	Different timing across municipalities of participation in ALMPs for persons on social assistance	No specific rules	+
Sweden	Hägglund (2011)	UI	12 months	2004	Random assignment	3 months	+
Norway/Sweden	Røed *et al.* (2002)	UI	14 months	1999–2000	Comparison of Norway and Sweden, different rules of UI system, differences in duration up until compulsory participation	Not mentioned	+

(Continued)

Table 1. *Continued*

Country	Author/Year	Program type	Duration until program start	Observation period	Identification strategy	Program duration	Threat effect
Australia	Richardson (2002)	UI	6 months	1997–1998	Introduction of MOI. Comparison of different age groups; individuals older than 24 years are not required to participate in ALMPs.	26 weeks to 2 years	NS
Switzerland	Lalive *et al.* (2008)	UI	Up to 17 months	1997–1999	Introduction of compulsory labor market programs, introducing differences in duration until compulsory participation across individuals	17 months	NS
Germany	Bergemann *et al.* (2011)	UI	Not mentioned	2007–2008	Self-reported probability of treatment and matching	Not mentioned	+
Finland	Tuomala (2011)	UI, social benefit, etc.	500 days	2003–2007	Introduction of compulsory labor market programs for persons receiving labor market support	Not mentioned	NS

Notes: A: Analyzes the same field experiment. B: Uses register data from the same or partly the same time period.
UI: Unemployment insurance. MOI: Mutual obligation initiative. The significance level reported in the threat effect column (far right) is a 5% significance level. NS: not statistically significant.

Germany, and the remaining 14 studies analyze programs in the Scandinavian countries (Denmark, Finland, Norway, and Sweden). Five of the nine Danish studies use registered data from the same or partially the same time period. Two of the nine Danish studies analyze the same field experiment. Thus, the 20 studies analyze a total of 15 unique populations. The excess of studies from Scandinavia may be due to these countries' historically high expenditure on ALMPs. To exemplify, Denmark, Sweden, and the USA spent 2.3%, 1.1%, and 0.1% of their respective GDP on ALMPs in 2011 (OECD, 2013).

Seven studies are based on randomized controlled trials (RCTs), while the remaining 13 studies are nonrandomized studies (NRSs). The USA, in particular, has a tradition of performing RCTs, and three of the six RCTs are American. Conversely, the Scandinavian tradition for collecting administrative data appears to facilitate NRSs. Only three of the 13 studies based on nonrandom assignment do not analyze Scandinavian systems.

3.1 *Identifying Strategies*

The identifying strategies concern the circumstances that enable researchers to identify a threat effect. If experimental data are available, the identifying strategy is random assignment of participants. If no random assignment is available, the effects are found using various identifying strategies as, for example, exploiting a "natural experiment" situation, where the threat effect can be identified by a policy change. To what extent the identification strategy used actually succeeds in identifying a threat effect involves assessment of selection bias. Concerning risk of bias, we therefore pay particular attention to selection bias.

In six instances, experimental data were available (analyzed in seven studies), where the individuals had been randomly assigned into a treatment group and a control group. Benus and Johnson (1997), Scrivener *et al.* (2002), and Black *et al.* (2003) investigate US programs, and because of the RCT design, the sample sizes remain relatively small. In another case, capacity constraints in the implementation of the ALMP led to random assignment of the individuals into a treatment group and a control group. Graversen and van Ours (2008) and Rosholm (2008) investigate the same field experiment, which was introduced in two Danish counties, and finally Pedersen *et al.* (2012) investigate another field experiment in Denmark. Hägglund (2011) uses experimental data from programs implemented in Sweden.

In Denmark and Finland, changes in the rules of ALMPs led to a "natural experiment" situation, where identification of the threat effect was made possible by changes in policy. The identification strategy was the exploitation of a shortening of the period in which individuals could remain unemployed before mandatory participation in an ALMP (Kyhl, 2001; Geerdsen, 2006; Geerdsen and Holm, 2007; Amilon, 2010; Tuomala, 2011; Clausen *et al.*, 2013).

Different comparison groups have also been applied in studies of the threat effect. Some of the ALMPs studied did not cover the entire labor force (i.e., not all unemployed are eligible for ALMPs). In this situation, it is possible to identify the threat effect by comparing individuals who were eligible for the program with individuals who were not eligible (Richardson, 2002; Lalive *et al.*, 2008). Differences in the timing of program participation between different regions are also used as an identifying strategy (Røed *et al.*, 2002; Graversen, 2004). A difference in self-reported perceived threat of participation in programs combined with matching was used in Bergemann *et al.* (2011). Finally, one study based the identification of the threat effect on functional form assumptions (Rosholm and Svarer, 2008) and another on exclusion restrictions (Jensen *et al.*, 2003). The identifying strategies of the included studies are listed in Table 1. Detailed characteristics of each of the 20 studies are available on request.

3.2 *Risk of Bias*

We assessed the methodological quality of the studies using an extension of the Cochrane Collaboration's risk of bias tool (Higgins and Green, 2011). Two reviewers independently assigned each selected study to the quality categories described below. All factors had to be scored either Met, Unclear, or Not met. Risk of bias assessment was based on two dimensions where selection bias was paid particular attention. The specific guidelines for scoring categories and the risk of bias assessment of the 20 studies are listed in Appendix D. Uncertainty or disagreement was solved by discussion with a third reviewer. Sensitivity analysis was carried out to evaluate whether the pooled effect sizes were robust across components of methodological quality, where possible.

3.2.1 *Selection Bias*

Random generation of allocation and allocation concealment were assessed for the RCT studies only (see Lachin, 1988; Schulz and Grimes, 2002a, b for the statistical properties of randomization and adequate methods of sequence generation and allocation concealment). Even if the RCT studies did not meet the criteria for random generation of allocation and allocation concealment (and rightly should be classified as quasi-randomized), we concluded that the threat effect was identified.

Concerning the NRSs, we have assessed two factors related to selection bias that only applies to the NRSs. The first factor is the extent to which the threat effect has been identified. Often, the threat effect is estimated by including a variable describing remaining time to benefit exhaustion in the model. This variable is a function of variables, which all have a direct effect on the individual's duration as unemployed, namely, initial entitlement, the duration of the unemployment spell and jumps in maximum duration realized while the spell goes on. Variation in remaining benefits will often come from one or more of these variables. Identification of a threat effect necessitates that at least one of these variables can be omitted from the modeling of the individuals' hazard out of unemployment. This, in turn, requires an assumption that either the variable does not have a separate effect on the individuals' hazard or that the effect follows a specific functional form, hence the term identifying assumption. We have investigated how the studies control for this kind of selection bias, that is, identifying assumptions used in the NRSs. We only conclude that the threat effect is identified if the authors clearly explain their identification strategy.

Even when appropriate identifying assumptions are made, there are a number of potential confounding factors that should be taken into account in an NRS. The second factor we have assessed for the NRSs is whether the authors as a minimum have controlled for labor market conditions, age, gender, education, and ethnicity.

Due to the nature of this intervention, special consideration regarding duration dependence is needed to avoid bias. This applies to both RCTs and NRSs. Most studies of unemployment find that the genuine duration dependence is negative, that is, the longer the unemployment spell, the smaller is the chance of finding a job due to, for example, loss of skills or self-confidence (see Serneels, 2002) for an overview. If the study does not disentangle the threat effect of ALMPs from the negative duration dependence, the estimated effect will be biased.

3.2.2 *Attrition Bias*

Special concern is needed about censoring. Due to the nature of the intervention, dropout cannot occur. The threat effect is often measured with survival data. Participants who do not leave the unemployment system before the end of the study are censored from the outcome data and if not adequately accounted for, it can introduce bias. Therefore, censoring of participants is a risk of bias, both in relation to the level of censoring and in relation to whether censoring is taken into account. We assessed whether censoring was less than or equal to 25% and taken into account.

4. Results

The 20 studies on the threat effect of ALPMs analyze a number of very different programs in various welfare regimes, and the results are not measured correspondingly. Due to difficulties in extracting a common outcome measure, it is not possible to combine all the studies' outcome measure in a meta-analysis. Two of the 20 studies do not provide data that permit the calculation of a standard error (Benus and Johnson, 1997; Black *et al.*, 2003). Two studies (Kyhl, 2001; Jensen *et al.*, 2003) are not included in the meta-analysis because we are concerned about whether they have succeeded to identify a threat effect (see Appendix B and the Supplementary Appendix). Of the remaining 16 studies, two studies analyze the same Danish field experiment employed in 2005–2006 (Graversen and van Ours, 2008; Rosholm, 2008). It was not possible to extract a comparable outcome measure in the study by Rosholm (2008).

Four NRSs analyze the same population using Danish register data from 1994 to 1998 (Geerdsen, 2006; Geerdsen and Holm, 2007; Amilon, 2010; Clausen *et al.*, 2013). We choose the study using an approach enabling us to pool most studies in a meta-analysis, who used a natural experiment to estimate individuals' hazard rate of leaving the UI system (Geerdsen, 2006).

Of the 12 studies potentially useful for inclusion in the meta-analysis, eight studies report comparable effect estimates that all uses indicator variables for the number of months or weeks leading up to compulsory participation in a program. The exact time points of measurement as well as country, year, duration of program, and duration of unemployment until program start for the studies are given in Table 2. Four of the studies include the point estimate 1 month into the active period, as program start occurs in practice with a delay.

4.1 *The Meta-Analysis*

We perform a meta-analysis based on the estimates from the eight studies presented in Table 2. Note that three of the studies are RCTs and five are NRSs. All eight studies use register data. Our decision to pool RCTs and NRSs can be considered as controversial: However, both methods have their pros and cons, and for the reasons given below, we believe it appropriate to pool the eight studies. Conceivably, RCTs have a low level of external validity compared to NRSs. However, the three RCT studies are reasonably large-field experiments, and we believe therefore that their results can be generalized to the same extent as the NRSs.

The shortcoming of NRSs concerns internal validity. Usually, the argument for not combining RCTs and NRSs concerns the inherent bias in NRSs opposed (implicitly assumed) to RCTs. This argument is only applicable if the RCTs are adequately randomized. Yet, it is clear that the allocation has not been properly randomized in the two of the three RCT studies included in the meta-analysis,[1] and should therefore be classified as quasi-experiments. The information concerning allocation method and concealment in the remaining RCT study included in the meta-analysis is very scarce.

Conversely, the authors of the NRSs identify the threat effect under careful consideration. All the NRSs except Røed *et al.* (2002) exploit changes in the rules of the ALMPs. This leads to a natural experiment situation, where the threat effect is identified by the policy change, proper identification assumptions, and econometric modeling. In addition, all of the NRSs correct for unobserved heterogeneity.

The three RCTs report a combined measure of the threat effect in the weeks preceding compulsory participation in a program, whereas the five NRSs report separate effect measures for the number of months or fortnights leading up to compulsory participation in a program. For those five studies, we calculate an average (synthetic) effect size to avoid dependence problems. This method provides an unbiased estimate of the average effect size but overestimates the standard error. We find evidence of heterogeneity, and we therefore rely on the results of random effects models. The use of random effects models when synthetic effect sizes are involved actually perform better in terms of standard errors than

Table 2. Characteristics of Studies Included in the Meta-Analysis.

Study	Country/year	Duration until program start	Time points of measurement	Program duration	Study design
Tuomala (2011)	Finland/ 2006	16–17months	Point estimates 1–2 and 0–1 months prior to program start were reported (average calculated by the review team)	Not mentioned	NRS
Pedersen *et al.* (2012)	Denmark/ 2008	0–13 weeks	A combined estimate of the effect the first 1—16 weeks of unemployment was reported	12 weeks	RCT
Lalive *et al.* (2008)	Switzerland/ 1997–1998	17 months	Point estimates 1 and −1 months prior to program start were reported (average calculated by the review team)	17 months	NRS
Richardson (2002)	Australia/ 1998–1999	6 months	Point estimates 3–4 fortnights and 1–2 fortnights prior to program start were reported (average calculated by the review team)	26 weeks–2 years	NRS
Geerdsen (2006)	Denmark/ 1994–1999	24–36 months	Point estimates 1, 0, and −1 months prior to program start were reported (average calculated by the review team)	75–80% of the time spend unemployed	NRS
Røed *et al.* (2002)	Sweden/ 1999–2000	14 months	Point estimates 1, 0, and −1 months prior to program start were reported (average calculated by the review team)	Not mentioned	NRS
Graversen *et al.* (2008)	Denmark/ 2005–2006	5–6 weeks and 4 months	A combined estimate of the effect 1+ weeks prior to program start was reported	3 months	RCT
Hägglund (2011)	Sweden/ 2004	12 months	A combined estimate of the effect on average 6.3 weeks from notification until program start was reported	3 months	RCT

Notes: RCT: randomized controlled trial. NRS: nonrandomized study. −1 month prior to program start: 1 month after program start.

fixed effects models, as it gives a more accurate estimate of the variance of the overall mean (Hedges, 2007).

Figure 1 shows the results from the meta-analysis of the eight studies. An insignificant negative threat effect is found in Tuomala (2011), while seven studies report results that indicate a positive threat effect; three of the study-level positive effects are statistically insignificant. The pooled results show a positive significant threat effect. The pooled estimate for the threat effect is a hazard ratio of 1.25; however, there

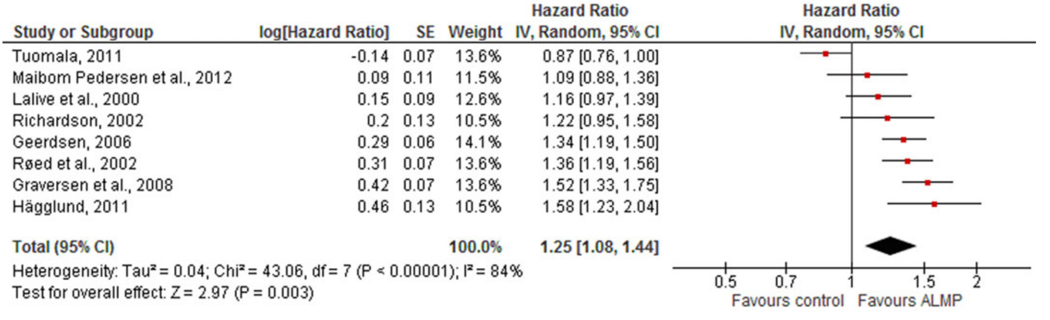

Figure 1. Threat Effect of ALMP (Hazard Ratio).

is significant heterogeneity of effects among studies (Q = 43.06, df = 7, p < 0.00001) but the point estimate is statistically different from zero (random effects 95% CI 1.08, 1.44, p = 0.003).[2]

Despite the power of random effects models is not great and therefore confidence intervals for pooled effects are fairly large; we find a statistically significant threat effect of ALMPs. Even though half of the studies included in the meta-analysis do not find a significant threat effect, the pooled estimate shows a significant positive threat effect, indicating the strength of the meta-analysis as it increases the statistical power.

4.2 *Sensitivity Analysis*

We perform three sensitivity analyses excluding studies that have a high risk of bias on the three factors: Consider baseline differences, disentangle duration dependence, and attrition bias (it is not possible to analyze the remaining components as there is no variation between studies, see Appendix B). The pooled estimate for the threat effect is a hazard ratio of 1.33 (random effects 95% CI 1.21, 1.45) when excluding studies that did not consider all baseline differences. Excluding studies that do not disentangle duration dependence yields a hazard ratio of 1.21 (random effects 95% CI 1.04, 1.41). Moreover, excluding studies that score Not met or Unclear on the attrition bias scale yields a hazard ratio of 1.39 (random effects 95% CI 1.24, 1.55). Thus, there were no noticeable changes in the results due to exclusion of studies with high risk of bias.

4.2.1 *Assessment of Publication Bias*

A scatterplot of individual treatment effects against standard errors (a Funnel plot) can be used for information about possible publication bias. A symmetric inverted funnel shape arises from a "well-behaved" data set, in which publication bias is unlikely, whereas an asymmetric funnel plot may indicate publication bias or could be caused by other factors (small study effects). The small number of studies that could be pooled, however, precludes the use of funnel plots (Higgins and Green, 2011). We can, however, not rule out the possibility of publication bias if not all studies carried out are published. The published studies may differ from the unpublished. Research with statistically significant results might be more likely to be published than work with nonsignificant results and perhaps also more likely to be published in higher impact journals. As we have searched for (and included) studies regardless of their publication status (i.e., working papers and other unpublished studies), we have minimized the risk of publication bias. The only risk of publication bias arises from the research that never leaves the researchers' desk.

4.3 *Findings of the Studies Not Included in the Meta-Analysis*

We briefly review the findings of the studies not used in the meta-analysis. Only one of the studies not included in the meta-analysis does not find any significant threat effect (Jensen *et al.*, 2003).

Four of the Danish studies not included in the meta-analysis analyze policy changes during the 1990s in Denmark (Geerdsen and Holm, 2007; Rosholm and Svarer, 2008; Amilon, 2010; Clausen *et al.*, 2013). All studies find significant threat effects. Geerdsen and Holm (2007) report that a 50 percentage point increase in the risk of enrolment in an ALMP results in approximately 50% increase in the hazard of leaving unemployment. Clausen *et al.* (2013) and Rosholm and Svarer (2008) find similar results. Amilon (2010) reports a significant threat effect on the number of search strategies, although modest in size.

Graversen (2004) reports a modest indirect threat effect (on duration dependence) based on municipality differences intends to activate early. Rosholm (2008) analyzes the field experiment also analyzed in Graversen and van Ours (2008). He finds a 20–40% increase in the exit rate from unemployment.

Three RCT studies from the USA all find a significant threat effect. Scrivener *et al.* (2002) analyze the threat effect for single parents in 1994–2001. They find a significant threat effect that increases when approaching ALMP. The findings in Black *et al.* (2003) look at a program targeted at individuals with the highest expected unemployment duration. Their findings suggest that threat of ALMP reduces unemployment duration by 2 weeks. Benus and Johnson (1997) find similar results in an experiment targeting all unemployed individuals. The threat of ALMP increases the exit rate by 5.7%.

Finally, Bergemann *et al.* (2011) find that the threat of ALMP increases job search in Germany.

5. Discussion and Conclusion

ALMPs have been one of the most highly advocated instruments for lowering high rates of unemployment, but they have also been heavily criticized for a lack of effectiveness, indicated by the several evaluations of the direct program effects (Heckman *et al.*, 1999; Gerfin and Lechner, 2002). If a large proportion of unemployed individuals choose to leave unemployment before they are required to participate in a compulsory ALMP, it may be difficult to avoid selection bias in evaluations of the ALMPs, as discussed by Heckman *et al.* (1999). Taking into account the fact that the threat effect may alter the evaluation of the total effects of a given program may be of great importance when the cost effectiveness of such programs is evaluated (Rosholm and Svarer, 2004a, b).

Another aspect of ALMPs is whether the existence of a threat effect is heavily dependent upon the success of public authorities to implement the system of ALMPs. Generally, most of the studies do not consider implementation issues, although a few of the studies briefly discuss the implementation of the ALMPs.

Moral hazard is often considered to be one of the major problems associated with benefit systems for unemployed workers. The findings of this review suggest that ALMPs are successful at reducing this problem. The main reason behind this effect appears to be that unemployed individuals speed up their job finding process (which might be a result of an increase in their search intensity or a lowering of their reservation wages when the time of the compulsory participation in an ALMP approaches).

In most cases, it appears that policy makers are not aware of the threat effects or that they are not openly recognizing them and for most ALMPs, the threat effect is not an explicit goal. In those cases, the unemployed individuals who are affected may actually be worse off than policy makers intend them to be, in the sense that they accept "bad" jobs to escape the imminent threat of program participation. Such detrimental side effects have not yet been fully investigated.

Our pooled results show a significant threat effect; a hazard ratio of 1.25 in the time period before ALMP participation (as indicated in Table 2). Although the effect is statistically significant, it is for practical use a modest effect. A more intuitive interpretation can be calculated as the probability of the threatened individuals finding a job first; that is, find a job before individuals who are not threatened by ALMPs:

HR = odds = $P/(1-P)$; $P = \text{HR}/(1+\text{HR})$ (Spotwood *et al.*, 2004). A hazard ratio of 1.25 corresponds to a 56% chance of the unemployed individual threatened with participation in an ALMP finding a job before an individual not threatened with ALMP participation. The lower and upper 95% confidence interval corresponds to 52%, respectively, 59% chance of finding a job first. A recent systematic review by Filges *et al.* (2015) that investigates the effect of participating in an ALMP estimates a hazard ratio of 1.09 corresponding to a 52% chance of finding a job before an individual not participating in an ALMP. This emphasizes the importance of recognizing the threat effect of ALMPs. Even though our pooled estimate shows a modest effect, it is larger than the actual effect of the programs. It is, however, lower than the "threat" effect of exhaustion of unemployment benefits. A recent review found an effect on the employment hazard ratio of 1.78 in the month of benefit exhaustion and 1.30 one month before benefit exhaustion (Filges *et al.*, 2013).

It is worth noting that we find a significant threat effect in a meta-analysis using studies of very different labor markets. The studies can be divided up according to the type of welfare regime: The universal welfare regimes (i.e., with provisions that cover everyone) in the Scandinavian countries, as well as Switzerland, have a comprehensive system of long-term programs, while the USA and Australia primarily employ short-term programs. If participation in ALMPs is viewed as a "tax" on leisure, then the universal regime, with its highly comprehensive programs, entails a major test of the unemployed. Eleven of the 20 studies are from Denmark, and the results of these studies are largely driven by the same institutional settings. However, in the meta-analysis, only three of the eight studies are Danish, and, in general, the results on the threat effect from the Danish studies do not differ from the results from other countries.

Some of the analyzed ALMPs were placed shortly after the start of a period of unemployment, while most of the programs started after 6 months or 1 year of unemployment. ALMPs placed at the start of the unemployment period involved a larger fraction of the group of unemployed than the programs starting after 1 year. Consequently, the early ALMPs affect many of those who would be able to find a job by themselves if they had to. The meta-analysis yields a significant threat effect when it includes not only those ALMPs placed shortly after the start of the period of unemployment, but also those ALMPs that did not start until after a year or more of unemployment. Consistently, the majority of the studies not included in the meta-analysis support this result, as they also find evidence of a threat effect independently of the timing of the ALMP (i.e., either placed shortly after the start of a period of unemployment or, for example, a year after).

Countries where individuals are able to stay unemployed for a long period of time before being forced to participate in an ALMP tend also to have programs of longer durations. This trade-off between the timing of compulsory ALMP participation and its duration may, in part, explain why a significant threat effect is found for both types of program designs: programs with duration of several months impose a higher "leisure tax" than programs lasting for 1 or 2 weeks only. The ALMPs analyzed are either early and short term or late and long term. Therefore, there is no evidence of whether short-term ALMPs placed later into the period of unemployment would entail a threat effect.

A limitation of our research is the limited amount of studies investigating the threat effect of ALMPs. We believe that we have reviewed all existing studies estimating the threat effect, but our total study population consists of only 20 studies, where only eight have comparable estimates for a meta-analysis.

With only eight studies, we cannot investigate heterogeneous effects and we are not able to do a thorough sensitivity analysis. Still, our findings are very clear; 15 of the 20 studies estimate a significant threat effect and the pooled estimate from the meta-analysis shows a hazard ratio of 1.25. That we are only able to find 20 studies from eight countries investigating that the threat effect clearly indicates that it is an area of research that needs to be explored further. If policy makers do not take the threat effect into account, they might underestimate the effect of ALMPs, as our findings suggest that although the effect is modest, the threat effect is larger than the effect of the actual program participation.

A major limitation of our research is that the overall quality of the evidence is not high. Few RCTs are conducted and none of them clearly demonstrate adequate randomization methods. Two of the RCTs should be classified as quasi-experiments as they use birth date to allocate participants. Overall, the threat effect of ALMP has not been evaluated with sufficient rigor. Well-designed RCTs are needed. Furthermore, two of the available studies did not provide data that permitted the calculation of a standard error. It is important that future studies not merely report on the statistical significance of their findings but also provide their results in sufficient detail to allow their inclusion in systematic reviews examining the magnitude of effects.

An interesting target for further research is the individuals that exit unemployment prior to the compulsory participation in programs and the exact time at which they exit. More information is needed about what triggers the transition out of unemployment: is the expected time of entering the program (according to legislation), or is the announcement of the ALMP's actual start date (in case the unemployed are unaware of the timing or the start date has been delayed) the trigger? Further research should also be directed at the possible side effects of the threat of ALMPs, in particular whether the individuals in question exit unemployment faster due to a higher acceptance of low-paid employment or whether some of them simply choose to withdraw from the labor force. These considerations point to the need for studies of the long-term consequences of exiting unemployment due to the threat of compulsory participation in an ALMP.

Acknowledgments

We thank Lars Pico Geerdsen, Niels Henning Bjørn, and Peter Jensen for valuable contributions to an earlier stage of this paper. We are grateful for all the comments and suggestions of the two anonymous referees of this journal.

Notes

1. Graversen *et al.* (2008) and Pedersen *et al.* (2012) used birth date to allocate.
2. The pooled fixed effects estimate yields a similar result: a hazard ratio of 1.24 (95% CI 1.18, 1.32). The weighted least-squares estimate is a hazard ratio of 1.24 (95% CI 1.10, 1.40). The inverse variance is used as weight.

References

Included Studies

Amilon, A. (2010) Active labour market programmes, job search and job finding in Denmark. *Labour* 24(3): 279–294.

Benus, J. and Johnson, R. (1997) Evaluation of the Maryland unemployment insurance work search demonstration. Working Paper for Maryland Department of Labor, Licensing and Regulation.

Bergemann, A., Caliendo, M., van den Berg, G.J. and Zimmermann, K.F. (2011) The threat effect of participation in active labor market programs on job search behavior of migrants in Germany. IFAU Working Paper 2011:4.

Black, D.A., Smith, J.A., Berger, M.C. and 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 Review* 93(4): 1313–1327.

Clausen, J.K., Geerdsen, L.P. and Tranæs, T. (2013) The motivation effect of active labor market policy on wages. *The Rockwool Foundation Research Unit*, Study Paper No. 50.

Geerdsen, L.P. (2006) Is there a threat effect of labour market programmes? A study of ALMP in the Danish UI system. *The Economic Journal* 116: 738–750.

Geerdsen, L. and Holm, A. (2007) Job-search incentives from labour market programs – an empirical analysis. *Labour Economics* 14: 639–652.
Graversen, B.K. (2004) The impact of active labour market programmes on welfare dependence in Denmark, Chapter 1 in PhD thesis 2004-2, Department of Economics, University of Aarhus, Denmark.
Graversen, B.K. and van Ours, J. (2008) How to help unemployed find jobs quickly: experimental evidence from a mandatory activation program. *Journal of Public Economics* 92: 2020–2035.
Hägglund, P. (2006) Are there pre-programme effects of Swedish active labour market policies? IFAU - working paper 2006:2.
Hägglund, P. (2011) Are there pre-programme effects of active placement efforts? Evidence from a social experiment. *Economics Letters* 112: 91–93.
Jensen, P., Rosholm, M. and Svarer, M. (2003) The response of youth unemployment to benefits, incentives, and sanctions. *European Journal of Political Economy* 19: 301–316.
Kyhl, T. (2001) Does the right and obligation to participate in "activation programmes" motivate the unemployed to look for work? *Nationaløkonomisk tidsskrift* 139(3): 278–297.
Lalive, R., van Ours, J. and Zweimüller, J. (2008) The impact of active labour market programmes on the duration of unemployment in Switzerland. *The Economic Journal* 118(525): 235–257.
Pedersen, J.M., Rosholm, M. and Svarer, M. (2012) Experimental evidence on the effects of early meeting and activation. IZA Discussion Paper 6970.
Richardson, L.L. (2002) Impact of the mutual obligation initiative on the exit behaviour of unemployment benefit recipients: the threat of additional activities. *The Economic Record* 78(243): 406–421.
Rosholm, M. (2008) Experimental evidence on the nature of the Danish employment miracle. IZA Discussion Paper 3620.
Rosholm, M. and Svarer, M. (2008) The threat effect of active labour market programmes. *Scandinavian Journal of Economics* 110(2): 385–401.
Røed, K., Jensen, P. and Thoursie, A. (2002) Unemployment duration, incentives and institutions – a micro econometric analysis based on Scandinavian data. Memorandum no. 9/2002, Department of Economics, University of Oslo.
Scrivener, S., Hendra, R., Redcross, C., Bloom, D., Michalopoulos, C. and Walter, J. (2002) Final report on Vermont's welfare restructuring project. Manpower Demonstration Research Corporation Report, available at www.mdrc.org.
Tuomala, J. (2011) The threat effect of mandatory programme participation in Finland - a study of activation reform of the labour market support system. *Labour* 25(4): 508–527.

Other References

Filges, T., Geerdsen, L.P., Knudsen, AS.D., Jørgensen, AM.K. and Kowalski, K. (2013) Unemployment benefit exhaustion: incentive effects on job finding rates: a systematic review. *Campbell Collaboration* 2013: 4.
Filges, T., Smedslund, G., Knudsen, AS.D. and Jørgensen, AM.K. (2015) Active labour market programme participation for unemployment insurance recipients: a Systematic Review. *Campbell Collaboration* 2015: 2.
Gerfin, M. and Lechner, M. (2002) A microeconometric evaluation of the active labour market policy in Switzerland. *The Economic Journal* 112: 854–893.
Hansen, C.T. and Tranæs, T. (1999) Optimal workfare in a society of workers and nonworkers. In T.M. Andersen and S.E.H. Jensen and O. Risager (eds.), *Macroeconomic Perspectives on the Danish Economy* (pp. 335–358). London: MacMillan Press.
Heckman, J.J., Lalonde, R.J. and Smith, J.A. (1999) The economics and econometrics of active labor market programs. *Handbook of Labor Economics* 3A: 1865–2097.
Hedges, L.V. (2007) Meta-analysis. In C.R. Rao (ed.), *The Handbook of Statistics* (pp. 919–953). Amsterdam: Elsevier.
Hedges, L.V. and Olkin, I. (1980) Vote-counting methods in research synthesis. *Psychological Bulletin* 93: 563–573.
Higgins, J.P.T. and Green, S. (eds.) (2011) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. Wiley-Blackwell and The Cochrane Collaboration. Available from www.cochrane-handbook.org.

Jackman, R. (1994) What can active labour market policy do? *Swedish Economic Policy Review* 1: 221–257.
Jenkins, S.P. (2005) Survival analysis. Lecture Notes, 18 July. http://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf
Lachin, J.M. (1988) Statistical properties of randomization in clinical trials. *Control Clin Trials* 9(4): 289–311.
Martin, J.P. and Grubb, D. (2001) What works and for whom: a review of OECD countries' experiences with active labour market policies. *Swedish Economic Policy Review* 8(2): 9–56.
Moffitt, R.A. (1996) The effect of employment and training programs on entry and exit from the welfare caseload. *Journal of Policy Analysis and Management* 15(1): 32–50.
OECD (2013) Labour market programmes: expenditure and participant. OECD Employment and Labour Market Statistics (database).
Rosholm, M. and Svarer, M. (2004*a*) *Estimating the threat effect of active labour market programmes*. Working Paper No. 2004–06, Department of Economics, University of Aarhus, Denmark.
Rosholm, M. and Svarer, M. (2004*b*) *Estimating the threat effect of active labour market programmes*. Discussion Paper No. 1300, IZA.
Schulz, K.F. and Grimes, D.A. (2002a) Generation of allocation sequences in randomized trials: chance, not choice. *The Lancet* 359(9): 515–519.
Schulz, K.F. and Grimes, D.A. (2002b) Allocation concealment in randomized trials: defending against deciphering. *The Lancet* 359(9): 614–618.
Serneels, P. (2002) Explaining non-negative duration dependence among the unemployed. CSAE Working Paper Series 2002-13, Centre for the Study of African Economies, University of Oxford.
Spotwood, L.S., Reid, J.E., Grace, M. and Samore, M. (2004) Hazard ratio in clinical trials. *Antimicrobial Agents Chemotherapy* 48(8): 2787–2792.
Stanley, T.D., Doucouliagos, H., Giles, M., Heckemeyer, J.H., Johnston, R.J., Laroche, P., Nelson, J.P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R.S. and Rost, K. (2013) Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys* 27(2): 390–394.
Van den Berg, G.J. (2001) Duration models: specification, identification, and multiple durations. In J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics, Vol. V (Chapter 55)*. Amsterdam: North-Holland.

Appendix

A *Search Methods for Identification of Studies*

A.1 *Electronic Sources*

Relevant studies have been identified through electronic searches of bibliographic databases, Internet search engines, etc. No language or date restrictions have been applied to the searches.

The search strategy used is presented below. It has been modified, when necessary, for the databases listed; in such cases, the full details of the modification have been reported. As NRSs are included in this review, trial filters have not been used.

The search was performed in the following electronic databases:

Econlit, IDEAS, ERIC, Psycinfo, InsídeConferences, Social Science Citation Index, Denmark's Research Database, and Danbib (Denmark's National Bibliography, including Library of Congress).

A.1.1 Search Terms.

1. Mutual adj obligations
2. (Active adj labo?r market adj policy)
3. lamp OR ALMP
4. (activation adj measure?)
5. (labo?r adj market adj program?)
6. search OR training OR (job adj training) OR seek? OR (job adj seek?)

7. education OR (class adj room adj teaching)
8. (workfare OR unemployment)/ti,ab)
9. 2–8/or
10. (incentive OR threat OR mutual) adj3 effect/ti,ab
11. 9 and 10
12. (entry adj effect?)/ti, ab
13. 12 and 10
14. 1 or 11 or 13

A.2 *Searching Other Resources*

Personal contacts with international researchers, developers, and independent investigators have been made to identify unpublished reports and ongoing studies. The references in reviews and primary studies have been scanned to identify new leads.

Google was used to search the web to identify potential unpublished studies. To search for European gray literature, we used OpenSIGLE (http://opensigle.inist.fr/). CEPR – Centre for Economic Policy Research (www.cepr.org) and NBER – National Bureau of Economic Research (www.nber.org) were searched to uncover potential discussion papers. Copies of relevant documents were made, recording the exact URL and date of access.

Relevant handbooks like *Handbook of Labour Economics, and Elsevier* have been searched to examine the references of their bibliographies.

B *Derivation of the Meta-Analysis*

The between study variance component, τ^2, is estimated using methods of moments. Random effects weighted mean effect sizes are calculated using 95% confidence intervals. The heterogeneity test statistic is given by:

$$Q = \sum\nolimits^{w_i} (\hat{\theta}_i - \hat{\theta}_{IV})^2,$$

where $\hat{\theta}_i$ is the individual effect estimates, $w_i = \frac{1}{(SE\{\hat{\theta}_i\})^2}$ and $\hat{\theta}_{IV} = \frac{\sum^{w_i} \hat{\theta}_i}{\sum^{w_i}}$.

Under the null hypothesis that there are no differences in threat effects among studies, this follows a chi-squared distribution with $1-k$ degrees of freedom (where k is the number of studies contributing to the meta-analysis). The I^2, which measures the extent of inconsistency among the studies' results, and is interpreted as the proportion of total variation in study estimates due to heterogeneity rather than sampling error, is calculated as

$$I^2 = max\left\{100\% \times \frac{Q-(k-1)}{Q},\ 0\right\}.$$

Each study's effect size is given the weight, $w_i' = \frac{1}{SE\{\hat{\theta}_i\}^2 + \hat{\tau}^2}$.

The summary effect size is given by

$$\hat{\theta} = \frac{\sum w_i' \hat{\theta}_i}{\sum w_i'}$$

$$SE\{\theta\} = \frac{1}{\sqrt{\sum w_i'}}.$$

C *Derivation of the Hazard Rate*

The hazard rate is defined as the event rate (in the present context, the event is leaving unemployment) at time t conditional on survival (staying unemployed) until time t or later. A hazard rate is constructed as follows (based on Jenkins, 2005; van den Berg, 2001):

The length of an unemployment spell for an unemployed individual is a realization of a continuous random variable, T. In continuous time, the hazard rate $\theta(t)$ is defined as

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

where the cumulative distribution function of T is $F(t) = \Pr(T < t)$ and the probability density function is $f(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(t \leq T < t+\Delta t)}{\Delta t} = \frac{dF(t)}{dt}$..$F(t)$ is also known in the survival analysis literature as the failure function and in the present context failure means leaving unemployment. $S(t)$ is the survivor function: $S(t) \equiv \Pr(T \geq t) = 1 - F(t)$; t is the elapsed time since the individual entered the unemployment system. Introducing covariates, the hazard rate becomes:

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

where $x(t, s)$ is a vector of personal characteristics that vary with unemployment duration (t) or with calendar time (s). The threat effect is the difference in hazard rates prior to compulsory participation between individuals who approach compulsory participation and individuals who do not approach compulsory participation. In all of the studies included in the meta-analyses, the threat effect is given as a proportional change in hazard rates. A proportional hazard rate is given by

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

where $\theta_0(t)$ is the baseline hazard, $\exp(x'\beta)$ is a scale function of the vector x of individual characteristics, and β is a vector of parameter estimates.

D *Assessment of Risk of Bias*

Selection bias

D.1 *Random Generation of Allocation*

Met (Resulting sequences are unpredictable [explicitly stated use of either computer-generated random numbers, table of random numbers, drawing lots or envelopes, coin tossing, shuffling cards, or throwing dice].)
Unclear (Vague statement that the study was randomized but not describing the generation of the allocation sequence.)
Not met (Explicit statement that the study was not randomized OR explicit description of inadequate generation of sequence (e.g., using case record numbers, alternation, date of admission, and date of birth.)

D.2 *Allocation Concealment*

Met (Participants and investigators cannot foresee assignment, for example, central randomization performed at site remote from trial location, sequentially numbered, sealed, opaque envelopes.)
Unclear (Vague statement that the study was randomized but not describing the concealment of the allocation sequence.)

Table A1. Risk of Bias Assessment.

Study	Benus and Johnson (1997)	Scrivener *et al.* (2002)	Black *et al.* (2003)	Jensen *et al.* (2003)	Rosholm (2008)	Graversen and van Ours (2008)	Hägglund (2011)	Richardson (2002)	Lalive *et al.* (2008)
Random generation of allocation	Unclear	Unclear	Not met	Not met	Not met	Not met	Unclear	Not met	NA
Allocation concealment	Unclear	Unclear	Not met	Not met	Not met	Not met	Unclear	Not met	NA
Identifying the incentive effect	Met	Met	Met	Not met	Met	Met	Met	Met	Met
Consider baseline differences in age, gender, education, ethnicity, and labor market conditions	NA	NA	NA	NA	NA	NA	NA	Met	Met
Disentangle incentive effect from duration dependence	Not met	Not met	Not met	Met	Met	Met	Not met	Met	Met
Attrition bias	Unclear	Unclear	Unclear	Met	Met	Met	Unclear	Met	Not met

(Continued)

Table A1. *Continued*

Study	Røed *et al.* (2002)	Kyhl (2001)	Geerdsen (2006)	Geerdsen and Holm (2007)	Rosholm and Svarer (2008)	Graversen (2004)	Tuomala (2011)	Amilon (2010)
Random generation of allocation	Not met	Not met	Not met	Not met	Not met	Not met	Not met	Not met
Allocation concealment	Not met	Not met	Not met	Not met	Not met	Not met	Not met	Not met
Identifying the incentive effect	Met	Not met	Met	Met	Met	Met	Met	Met
Consider baseline differences in age, gender, education, ethnicity, and labor market conditions	Met	Met	Met	Met	Met	Not met	Not met	Met
Disentangle incentive effect from duration dependence	Met	Not met	Met	Met	Met	Met	Met	Met
Attrition bias	Unclear	Not met	Met	Met	Unclear	Met	Unclear	Unclear

Study	Clausen *et al.* (2013)	Bergemann *et al.* (2011)	Pedersen *et al.* (2012)
Random generation of allocation	Not met	Not met	Not met
Allocation concealment	Not met	Not met	Not met
Identifying the incentive effect	Met	Met	Met
Consider baseline differences in age, gender, education, ethnicity, and labor market conditions	Met	Met	NA
Disentangle incentive effect from duration dependence	Met	Met	Met
Attrition bias	Met	Met	Unclear

Note: NA: Not applicable.

Not met (Explicit statement that allocation was not concealed OR statement indicating that participants or investigators can foresee upcoming assignment [e.g., open allocation schedule, unsealed, or nonopaque envelopes].)

D.3 *Identifying the Threat Effect*

Met (Clearly explained how the study identifies an threat effect.)
Unclear (Sufficient information could not be obtained.)
Not met (No explanation or statement about identification.)

D.4 *As a Minimum Take into Account Baseline Differences in Age, Gender, Education, Ethnicity, and Labor Market Conditions*

Met (Clearly stated that all confounders are taken into account or there was no initial imbalance.)
Unclear (Sufficient information could not be obtained.)
Not met (Initial imbalances and no statement that all confounders have been taken into account.)

D.5 *Disentangle Threat Effect from Duration Dependence*

Met (Clearly stated that the threat effect has been disentangled from the duration dependence or the authors clearly state that they assume duration dependence is not present.)
Unclear (Sufficient information could not be obtained.)
Not met (No statement about duration dependence.)

D.6 *Attrition Bias*

Met (Censoring less than or equal to 25% and taken into account.)
Unclear (Censoring not reported.)
Not met (Censoring greater than 25% or not taken into account.)