

Evaluating active labour market policies using randomised control trials

Johan Vikström

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Evaluating active labour market policies using randomised control trials^a

by

Johan Vikström^b

February 10, 2026

Abstract

This paper discusses randomised controlled trials (RCTs) of labour market policies in the Nordic countries. Evaluating labour market policies is essential to ensure that public resources are allocated to effective interventions that improve employment outcomes, yet credible evaluation is challenging because programme participation is often selective. RCTs address this challenge by creating comparable treatment and control groups and are therefore widely regarded as the gold standard in policy evaluation. Using illustrative examples, the paper illustrates key issues related to motivating, designing, implementing, and analysing RCTs in a labour-market policy context. The examples cover interventions such as job-search assistance, monitoring, private service provision, and light-touch behavioural policies. The paper discusses ethical considerations, randomisation designs, implementation challenges, and analytical issues including take-up, outcome measurement, cost-effectiveness, and general equilibrium effects. It concludes that while the existing evidence base is substantial, continued experimentation and causal evidence remain important as labour market institutions and policies evolve. Building institutions that support RCTs is identified as one challenge.

^a The paper benefited from comments from Anders Forslund, Kari Hämäläinen, Antti Kauhanen, Martin Lundin, Jukka Mattila, Roope Uusitalo, and of participants at research seminars of Nordic Policy Economic Review.

^b IFAU, Uppsala universitet och UCLS, epost: johan.vikstrom@ifau.uu.se

Table of contents

1	Introduction.....	3
2	Three examples of Nordic RCTs	4
2.1	Private providers in Denmark	6
2.2	Early meetings in Sweden.....	8
2.3	Light-touch intervention in Norway	11
3	Motivating, conducting and implementing RCTs.....	12
3.1	Motivating RCTs in the context of labour market policies.....	12
3.2	Ethics	14
3.3	Randomisation designs	15
3.4	Pre-analysis plans.....	16
3.5	Implementation challenges	16
3.6	Building institutions that support RCTs.....	17
4	Analysing RCTs of labour market policies.....	19
4.1	Take up	19
4.2	Objective outcomes.....	20
4.3	Costs.....	21
4.4	Nordic data and analyses of mechanisms	23
4.5	General equilibrium effects.....	24
5	Concluding remarks	25
	References	27

1 Introduction

Evaluating labour market policies is essential to ensure that public resources are used effectively and that policies achieve their intended goals. Labor market interventions—such as training programmes, job-search assistance, or wage subsidies—are often costly and can have far-reaching effects on individuals, employers, and the broader economy. Without systematic evaluation, policymakers risk keeping ineffective or even harmful programmes going, while potentially overlooking initiatives that deliver substantial benefits. Rigorous evaluation provides credible evidence on what works, for whom, and under what conditions, enabling more efficient policy design and better targeting of support. In addition, evaluation can help uncover the mechanisms behind observed outcomes, offering insights into why certain programmes succeed or fail.

Conducting credible evaluations is challenging, however, because it is often difficult to find a credible comparison group. A non-experimental evaluation typically compares job prospects for unemployed job seekers who participate in a labour market programme or intervention with those of a comparison group consisting of unemployed job seekers who do not participate. The issue is that job seekers and caseworkers make choices based on individual circumstances, motivation, preferences, and constraints, which can lead to bias in comparisons between participants and non-participants. As a result, those who end up participating in a policy intervention may differ systematically from those who do not, and adjusting for these differences between the treatment and control groups may be difficult or even impossible. As a result, non-experimental evaluations struggle to distinguish between the programme's true effects and differences that simply reflect who participates in it.

Randomised controlled trials (RCTs) provide a powerful way to overcome these challenges. By randomly assigning individuals, firms, or regions to treatment and control groups, RCTs create comparable treatment and control groups, enabling credible causal inference. They eliminate selection bias due to self-selection from the policy intervention. In doing so, RCTs generate credible, actionable evidence that can guide better policy design, improve accountability, and ensure that public funds are spent on initiatives that make a real difference in people's employment prospects.

Even if RCTs are the gold standard in evaluations, there are challenges and opportunities related to motivating, designing, implementing, and evaluating RCTs in a labour-market policy context. This paper uses three examples of RCTs conducted in the Nordic countries to discuss these topics and to present key empirical findings from Nordic RCTs. It does not provide a full review of the existing experimental and non-experimental evidence, but Section 2 starts with a brief overview of Nordic RCTs. Section 2 also introduces the three main RCT examples referred

to throughout the paper, which serve as background for the subsequent sections. These include a Danish RCT of private providers, a Swedish RCT of early meetings, and a Norwegian RCT featuring a light-touch intervention aimed at promoting good habits.

Section 3 then analyses topics related to motivating, conducting, and implementing RCTs. It discusses the main advantages of RCTs in a labour market context and explains why they are considered the gold standard for understanding which labour market policies work and for whom. This section also covers ethical considerations when conducting RCTs, different randomisation designs, pre-analysis plans, implementation challenges, and research independence. One conclusion is that RCTs are often ethical to conduct. When difficult allocation decisions must be made, and resources are limited, random assignment can be as ethical as any other method for assigning job seekers to an intervention. This section also touches on several practical aspects of implementing RCTs and concludes that careful implementation is key.

Section 4 focuses on topics related to analysing RCTs. Even though RCTs are designed to allow simple comparisons between randomised treatment and control groups, several important factors should be considered when conducting and analysing them. These include issues related to take-up and the importance of using objective outcomes. In addition, to using objective outcomes, it is also important to measure the costs of policy interventions, as cost-effectiveness is often a central concern when assessing outcomes. Section 4 also illustrates how the Nordic countries are well endowed with excellent administrative register data and how this can be leveraged in RCTs. A complicating factor when conducting RCTs is the presence of general equilibrium (GE) effects—that is, when support for some job seekers may displace employment or affect non-treated job seekers in various ways. Such responses and their implications for RCTs are also discussed in Section 4.

2 Examples of Nordic RCTs

Over the years, several RCTs on labour market policies have been conducted in the Nordic countries, most notably in Sweden and Denmark. This section begins with a brief overview of key Nordic experiments without providing an exhaustive review of all RCTs. It then discusses three examples on three separate research topics from three Nordic countries. These examples are then used to illustrate key aspects of RCTs in the subsequent sections.

2.1 A brief overview

A very early experiment on job-search assistance was conducted as early as 1975 in Eskilstuna, Sweden, and included 500 long-term unemployed individuals¹. The intervention targeted both employers—by informing them about job seekers, influencing their labour demand specifications, and enabling subsidized employment—and job seekers—by improving and broadening their job search and assisting with employer contacts. Overall, the intervention had positive employment effects.

After this early exception, a wave of experiments has been carried out in Finland, Denmark, and Sweden since the 1990s. In 1996–97, an RCT of a week-long job-search training workshop was conducted in Finland. The experiment was inspired by a previous RCT conducted in the US (see Vinokur et al., 2000). The aim was to change the role of participants from passive unemployed individuals to active job seekers as a means of helping them find jobs again. The evaluation conducted by Vuori et al. (2002) reveals a positive effect on self-reported job stability but no effect on employment rates or earnings.

This was followed by RCTs conducted in Sweden in 2004 (Hägglund, 2011, 2014) and in Denmark in 2005–06 (Graversen and van Ours, 2008a). For the Swedish RCTs, Hägglund (2014) compares the relative efficiency of combining job-search monitoring and job-search assistance with monitoring alone, while Hägglund (2011) estimates the effects of an invitation to a meeting aimed at monitoring search activity and assisting with more effective job searches. The Danish RCT (called *Hurtig i gang*) was implemented in two counties and consisted of early and mandatory participation in a job-search assistance course, frequent meetings with employment officers, and programme participation after a few months of unemployment. An initial evaluation by Graversen and van Ours (2008a) found large positive effects on employment.²

Since those early experiments, a variety of RCTs have been conducted in the Nordic countries. Broadly speaking, Denmark was an early mover, with several experiments between 2005 and 2011 including the *Hurtig i gang* RCT mentioned above, the follow-up RCTs (*Hurtig i gang 2*) evaluated by Rosholm and Svarer (2009a), the youth intervention (*Unge – godt i gang*) evaluated by Høeberg et al. (2011), an intervention targeting long-term unemployed people (*Alle i gang*) evaluated by Rosholm and Svarer (2009b), as well as experiments on early meetings (Maibom et al., 2017) and outsourcing labour market policies to private-sector providers (Rehwalder et al., 2017).

More recently, Sweden has been particularly active in conducting RCT evaluations of a wide range of topics, including Engström et al. (2012), Bennmarker et al. (2013), Laun et al. (2014),

¹ The RCT is described in the Swedish government report SOU 1978:60.

² Subsequent evaluations of the same experiment include Graversen and van Ours (2008b), Rosholm (2008), Blasco and Rosholm (2011), Vikström et al. (2013), Sørensen (2016, 2017) and Gautier et al. (2018).

Helgesson et al. (2020), Fogelgren et al. (2023), Egebark et al. (2024), Dahlberg et al. (2024), Cheung et al. (2025), Hensvik et al. (2025), and Cockx et al. (2025). RCTs in Norway include Bjorvatn et al. (2021), Berg et al. (2021), Sveinsdottir et al. (2021), and Hernæs (2025). In Finland, they include Malmberg-Heimonen (2005) and Pesola et al. (2025). Note that this is not an exhaustive list of all RCTs; in particular, it excludes some interventions targeted at individuals with mental health issues and interventions related to disability insurance that are less directly connected to labour market policy.

These RCTs span a range of different policies, including evaluations of early and frequent meetings with caseworkers and other job-search initiatives (Hägglund, 2014; Cheung et al., 2025), activations programmes (Graversen and van Ours, 2008ab; Rosholm, 2008; Vikström et al., 2013; Gautier et al., 2018), group and remote meetings (Maibom et al., 2017), private providers (Bennmarker et al., 2013; Laun et al., 2014; Egebark et al., 2024; Hernæs, 2025), online job recommendations (Hensvik et al., 2025), light-touch interventions with behavioural aspects (Bjorvatn et al., 2021; Cockx et al., 2025), supported employment (Berg et al., 2021; Fogelgren et al., 2023), early intervention programmes for immigrants (Sveinsdottir et al., 2021; Dahlberg et al., 2024), other immigration integration services (Pesola et al., 2025), intensive job-search assistance for newly arrived immigrants (Joonas and Nekby, 2012; Helgesson et al., 2020), vacancy referrals (Engström et al., 2012), intensified monitoring (Hägglund, 2014), displacement effects (Gautier et al., 2018; Cheung et al., 2025), information about rules and regulations (Cairo and Mahlstedt, 2023), policies that combine monitoring and job-search assistance (Hägglund, 2011) as well as interventions that combine job-search support and psychological aspects (e.g., Vuori et al, 2002).

Even if the current evaluations cover a wide range of interesting topics, more evidence is needed to support efficient policy design. New policies are frequently introduced, and existing ones are often reformed, which may alter their effectiveness. More evidence is also needed on which programmes work best for which groups. In some cases, it may be difficult to translate results from one country to another due to institutional and labour market differences. These are just some of the many reasons why additional RCTs remain valuable, despite the growing body of evidence from previous Nordic studies.

2.2 Private providers in Denmark

I now take a closer look at three RCTs, starting with providers in Denmark. Historically, employment services have been provided by government or municipal job-centers. In recent decades, however, many countries have moved toward and/or experimented with contracting out various employment services to private providers, e.g., training courses and job-search assistance,

either giving private providers full responsibility for all employment services provided to the job seeker or contracting out only certain parts (e.g., a single training course).

To study the effects of private providers, Rehwald et al. (2017) evaluate a Danish RCT which compares private and public employment services. One underlying motivation is that, in theory, it is ambiguous whether public or private employment services are more effective. On one hand, competition among private providers and financial incentives may spur innovation (e.g., novel ways to help job seekers) and cost savings (e.g., less bureaucracy and greater resources spent helping the job seekers). On the other hand, contracting out involves transaction costs for setting up, monitoring, and regulating markets. It may also lead to unwanted cost cuts if private providers generate profits by lowering the quality of services (e.g., fewer meetings) provided to job seekers. Thus, can market provision deliver on its promises, or should employment services be provided in-house by public employment agencies?

The RCT evaluated by Rehwald et al. (2017) was conducted in 2011, with newly registered unemployed university graduates as the target group. The background is that, in Denmark, the municipalities are responsible for organising employment services, and in 2011, contracting out services to private providers was the default option for unemployed academics. That is, the municipalities were required to contract out services. The RCT included 3,107 job seekers in four municipalities. The treatment group received the default option to go to a private provider, and the control group received support directly from the municipal job centres.

Private-provider systems may be organised in different ways, and their efficiency may depend on the exact design. In Denmark, 17 private providers were selected through a tendering round, where interested providers submitted offers, including a bid price for the total cost per job seeker per year. The public agencies then assigned job seekers to these selected providers, and the providers were not allowed to refuse any job seekers. The private providers were responsible for helping job seekers return to work using an employment-oriented strategy. Given basic requirements regarding the number of meetings and programme participation, the providers were free to organise the support as they saw fit. The compensation was heavily focused on a performance-related bonus (75% of the total compensation and 25% as basic compensation), awarded if the job seeker obtained regular employment for at least 13 weeks.

The findings from Rehwald et al. (2017) can be summarised as follows. On average, there are no differences in labour market outcomes (regular employment, subsidised employment, benefit receipt, or unemployment) between participants receiving public and private services. However, private provision is significantly more expensive, making it less cost-effective from a public spending perspective. Without transfer payments, the costs were DKK 10,392 for job seekers assigned to private services, while the costs for the public option were DKK 2,144 lower, i.e.,

21% lower. In terms of strategies, private providers held more frequent meetings and delivered more intensive, employment-oriented, and earlier interventions. By contrast, client satisfaction was reported to be slightly higher among jobseekers assigned to the public provider.

These conclusions align with results from several other RCTs of private provision. Focusing on RCTs from the Nordic countries, Bennmarker et al. (2013) analyse the effects of contracting out placement services for hard-to-place jobseekers in Sweden. They find that private providers generally do not improve labour market outcomes in terms of employment rates, earnings, and months worked. Unfortunately, they are not able to study costs. In a slightly different setting, Laun and Thoursie (2014) examine the effects of vocational rehabilitation for the long-term sick and temporarily disabled in Sweden and find no difference in labour-market outcomes or costs between private contractors and public services. More recently, also in Sweden, Egebark et al. (2024) evaluate a large-scale RCT of various aspects of private providers. One result is that the private providers were substantially more expensive (46–60% more, depending on the public sector support with which it is compared), but they did not generate more employment or higher labour earnings. They also show that private providers that are paid more do not generate better employment results.

Compared to the other studies, Rehwald et al. (2017) distinguish themselves in several ways. It provides the first Danish RCT evidence on contracting out employment services. It focuses on academics, which is relevant, as they may face a different labour market from other job seekers. It also evaluates a case involving the outsourcing of the entire package of employment services. The authors are able to leverage exceptionally rich data, such as reliable and frequent administrative data on employment, as well as detailed data on costs. The latter is key, as the merits of private providers largely hinge on the cost side, given that private providers do not improve employment outcomes.

2.3 Early meetings in Sweden

Evidence suggests that early meetings and other types of job search assistance, aimed at helping job seekers search for jobs more efficiently, is one of the more powerful labour market policies (Card et al. 2010, 2018). This evidence includes results of several Nordic RCTs (see, e.g., Maibom et al., 2017 for a discussion). Early meetings may have positive direct effects on those treated. However, they may also lead to substantial displacement of jobs for the non-treated. In particular, if early meetings or any other labour market policy help some treated job seekers find jobs, this may come at the expense of other job seekers searching for jobs in the same local labour market. That is, ALMPs may lead to displacement effects, whereby jobs that non-treated individuals would otherwise have obtained are taken by treated participants. This is crucial from a policy perspective as displacement mainly represents a reordering of job queues rather than net

employment gains. However, compared to the direct effects of early meetings, credible evidence on such displacement effects is scarce. Exceptions, all conducted outside the Nordic countries, include Crépon et al. (2013), Gautier et al. (2018), and Ferracci et al. (2014).

This type of displacement of jobs represents one aspect of general equilibrium (GE) effects. Other GE responses may include changes in wage bargaining, matching efficiency, or firm behaviour—such as vacancy creation and market entry. For example, if an intervention helps job seekers search more efficiently, firms may fill vacancies faster, making it more profitable to post additional vacancies and thereby reducing unemployment. Conversely, if a policy strengthens workers' bargaining positions (e.g., by improving outside options), wages may rise, reducing firm profitability and potentially lowering employment. Such GE responses are highly relevant to consider when evaluating labour market policies. For instance, imagine a policy intervention that helps treated job seekers find jobs faster, leading to the conclusion that the policy is effective. However, if this policy generates a high degree of job displacement or distorts wage bargaining, these negative GE employment responses could outweigh the positive employment effects for treated job seekers—ultimately reversing the conclusion about the policy's effectiveness.

Considering the importance of displacement effects and the scarce evidence, Cheung et al. (2025) implemented and evaluated a Swedish RCT specifically designed to capture displacement of jobs. The policy intervention consisted of three extra meetings with a caseworker during the first quarter of unemployment for all newly unemployed job seekers. To capture job displacement, these early meetings were evaluated using a two-level randomisation design, with randomisation over both job seekers within local offices and across local offices. The within-office randomisation of job seekers identifies whether the treated job seekers are more likely to find a job (the direct effect). This is similar to a standard RCT design with a randomised treatment and control group. The across-office randomisation identifies displacement and answers to what extent this comes at the expense of non-treated job seekers in the same labour market. Essentially, this part of the analysis compares non-treated job seekers at offices where half of the job seekers are randomised to the treatment (active offices) with non-treated job seekers at offices without any treated (non-active offices). Since neither group is called to early meetings, this difference measures the job-displacement effect arising from increased competition from treated job seekers in the same office.

The study documents a positive direct effect of early meetings on the re-employment rate by about 4.6%. However, it also finds substantial displacement, as the re-employment rate for the non-treated job seekers decreases by 3.8%. The implication is that failing to account for the job displacement would overstate the average impact by 89%. Nevertheless, the effect on overall employment, taking both direct and displacement effects into account, is still positive. The

substantial displacement effects are consistent with previous studies conducted outside the Nordic countries (Crépon et al. 2013; Gautier et al. 2018). Face-to-face and distance meetings have equally positive employment effects, supporting the case for using communication technologies to provide job search assistance.

Cheung et al. (2025) also study the mechanisms behind these effects. They find that early meetings are effective because they increase the number of vacancy referrals passed from caseworkers to job seekers. Using their expertise, caseworkers are able to direct job seekers to relevant job openings, which helps job seekers apply to the most suitable positions at an earlier stage. By examining resource allocation, Cheung et al. (2025) also conclude that the observed displacement effects are due to job displacement rather than resources being reallocated from the control group to participants. Another result is that the effects of early meetings are more positive under favourable labour market conditions than in weaker labour markets with high unemployment rates.

Cheung et al. (2025) also shed more light on the GE effects. They analyse vacancy data and show that vacancies increase more in the active areas with the early meetings than in the control areas without any additional early meetings. The idea being that the improved job search as a result of the early meetings helps firms fill their open vacancies faster, making it more profitable to open more vacancies. Such a GE response contributes to the positive effects of the early meetings. Interestingly, Cheung et al. (2025) also find evidence of a delayed vacancy response, with larger changes in the vacancy rate toward the end of the experiment. That is, it takes time for firms to observe and react to the changes in job seeking caused by the early meetings. It suggests that job-search assistance may have more positive effects in the long term because, in the short term, the programme mainly leads to search congestion, while in the long term, improved searches induce firms to open more vacancies, increasing job finding rates and pushing down the unemployment rate.

To connect the different results, Cheung et al. (2025) combine the RCT variation with a structural analysis. The basic idea is to build a model that incorporates the improved job search, the job displacement, firms' creation of vacancies and any wage impacts. Without going into details, it involves using the estimates from the RCT to estimate a structural model and then use the estimated model for policy simulations. One simulation considers the employment effects of a full-scale rollout of the early meetings to all newly unemployed job seekers in Sweden. The simulations show that increasing the share of treated from 0% to 100% lowers the unemployment rate by around 0.2 percentage points, suggesting that the net effect of a full-scale rollout is positive despite the substantial displacement. The costs also decrease. When taking the observed delayed vacancy response into account and simulating a longer-term response to the early meetings, the

overall assessment of early meetings becomes even more positive. This type of structural analysis has become more popular in recent years, and combining RCTs and structural analyses should be a key area for future research in the Nordic countries.

2.4 Light-touch intervention in Norway

The Norwegian RCT evaluated by Bjørvatn et al. (2021) begins with the idea that unemployment is a complex problem with many potential causes and no single solution. Traditionally, research has emphasised structural explanations such as inadequate skills, perverse incentives, and discrimination. However, a growing body of literature in psychology and behavioural economics suggests that behavioural factors, such as low motivation and bad habits, can lower job-finding rates among unemployed job seekers. There are several potential reasons for this. Unemployment may disrupt daily routines and lead to the loss of social connections, which in turn may foster bad habits and negatively affect health and psychological well-being. Altogether, this may hinder an efficient job search.

Building on this background, Bjørvatn et al. (2021) designed and tested a light-touch goal-setting intervention aimed at counteracting harmful habits. The starting point was that the Norwegian public agency NAV identified a target population of 2,848 individuals aged 16–29 years in one province who had been unemployed for at least eight weeks. They received an SMS invitation to participate in a study, and the analysis sample consists of the 684 individuals who completed a baseline survey embedded in the experiment. It consisted of a set of background questions, and the goal-setting intervention evaluated in the RCT. This treatment was only given to the treatment group (two-thirds of the sample), and half of these participants also received SMS reminders about the goals. The remaining one-third in the control group only received the initial background questions in the survey.

Recognising that lifestyle improvements may play a crucial role in facilitating re-employment, the intervention focused on three “keystone” habits: sleep, exercise, and substance use. The underlying premise was that explicitly setting goals for healthier daily routines could enhance participants’ prospects of finding employment. Specifically, it was divided into three steps: 1) questions about the job seekers’ current situation; 2) asking participants to reflect on their current habits and express goals for the future; and 3) a summary of the expressed goals, along with encouragement to take a screenshot and remember them. The SMS reminder component of the intervention consisted of short weekly messages reminding participants of their goals. To evaluate the effects, the authors used both labour market outcomes based on administrative register data and follow-up surveys, focusing, for instance, on habits.

Fascinatingly, this very low-cost goal-setting intervention led to improved labour market outcomes. Twelve months after the intervention, the probability of being employed was 7

percentage points higher, and the share of participants receiving unemployment insurance benefits was 5 percentage points lower. This clearly shows that the intervention boosted employment rather than causing exits from the labour force. Moreover, about half of the participants reported wanting to exercise more and go to bed earlier, and these habits did improve following the intervention. Participants also demonstrated a more positive mindset and improved overall life satisfaction. There are also indications of an increased number of job applications. All this resulted from an essentially costless intervention without any apparent distortive effects on those already employed. The interpretation is that the act of setting goals itself may have generated a more optimistic outlook, reinforcing the changes in habits and contributing to improved labour market outcomes.

Bjørvatn et al. (2021) add to the growing literature on light-touch interventions. In terms of Nordic RCTs, one example is Cockx et al. (2025), who designed and evaluated an RCT of motivational emails designed to motivate job seekers to continue to actively search for jobs. Hensvik et al. (2025) study how AI can alleviate search friction and help job seekers search for jobs online. Using an RCT, they evaluate an online job-recommender system that uses job seekers' click history to generate relevant personalised job recommendations. This study also relates to other non-Nordic RCTs. Two notable studies are Altmann et al. (2018) on the use of informational job-search folders and Belot et al. (2019) on the impact of low-cost online job-search advice.

3 Motivating, conducting and implementing RCTs

It is challenging to motivate, prepare and implement RCTs, and multiple pitfalls may ruin the trial. Some challenges relate to ethics, preparing pre-analysis plans, management of the trial, adjusting to unforeseen circumstances and questions related to research independence. This section discusses some implementation topics.

3.1 Motivating RCTs in the context of labour market policies

Designing effective labour market policies requires reliable evidence on what works, for whom, and under what conditions. Yet policy interventions are often complex, costly, and implemented in environments influenced by many confounding factors, making it difficult to identify their true impact. One key challenge is selection into a particular intervention, policy, or programme. A traditional policy evaluation based on non-experimental data compares the outcomes of treated participants with those of a comparison group, perhaps adjusting for observed background characteristics. This approach requires that the treatment and control groups are comparable in the absence of treatment. However, in practice, job seekers' and caseworkers' choices, actions,

and preferences often mean that certain types of job seekers are more likely to be enrolled in certain policies. For instance, one policy may attract highly skilled and motivated job seekers, while another may attract those further removed from the labour market. Thus, without properly controlling for selection into the treatment group, comparisons between the treatment and control groups may lead to highly misleading conclusions, as such comparisons reflect both the selection and the policy and impacts. As a result, many non-experimental evaluations may capture selection effects rather than true causal impacts.

RCTs offer a powerful solution to these challenges. By randomly assigning individuals, firms, or regions to treatment and control groups, RCTs ensure that, on average, the treatment and control groups are comparable in terms of both observable and unobservable characteristics. This allows researchers to isolate the causal effect of an intervention, free from selection bias and other challenges (e.g., reverse causality). In other words, the major advantage of RCTs is that they help resolve selection problems and provide credible answers to policy-relevant questions. To illustrate, consider the three RCTs above.

Rehwald et al. (2017) evaluate an RCT comparing public and private provision of ALMPs—a high-stakes policy question often shaped by ideological positions. Without randomisation, such evaluations typically rely on comparisons between job seekers who choose or are assigned to private providers and those who remain with public ones. Even after adjusting for observable characteristics, unobserved factors such as motivation or preferences could bias the results. If more motivated job seekers select private providers, the estimated effect would be positively biased; if less employable individuals are assigned to private providers, the bias would be negative. In either case, non-randomised comparisons risk producing misleading conclusions. This is shown explicitly by Bennmarker et al. (2013), who report that a non-experimental evaluation of private providers indicates a positive employment effect, whereas their RCT shows that there is no positive employment effect of private providers.

Cheung et al. (2025) evaluate frequent early meetings. In a non-randomised setting, job seekers called to more meetings may differ systematically from those called to fewer. Public employment agencies often allocate more meetings to individuals who need extra support, while stronger job seekers—who might find work quickly—receive fewer. As a result, simple comparisons could underestimate the real benefits of frequent meetings, since the control group would include stronger job seekers more likely to succeed regardless of support. Third, consider the novel intervention in Bjørvatn et al. (2021), which encourages job seekers to set personal goals related to daily habits such as sleep, exercise, and substance use. In a non-randomised implementation, agencies might target job seekers with poor baseline habits with such an intervention—a group

likely to have weaker labour market attachment. This would then bias the estimated impact downward, underestimating the intervention's true effect.

In all these cases, explicit randomisation is essential to uncover the true causal impact and to avoid incorrect policy conclusions. While some non-experimental methods—such as exploiting rule changes, policy discontinuities, exogenous variation, or rich administrative data—can produce credible estimates, RCTs remain the gold standard, offering the most robust causal evidence. The importance of RCTs may differ, but generally, the benefits are greatest when selection into treatment is particularly challenging to address non-experimentally, there is limited prior credible evidence on the policy, and/or the policy is new, and random assignment is a fair and feasible allocation method.

Beyond estimating average effects, RCTs can also identify heterogeneous impacts, revealing which groups benefit most and informing policy refinement. RCTs also foster policy learning and innovation, allowing decision-makers to test novel interventions—such as light-touch or behavioural policies—on a small scale before wider rollout. In short, RCTs bring scientific rigour to labour market policymaking, enabling governments to move beyond assumptions and ideology toward evidence-based interventions grounded in proper causal evaluation.

3.2 Ethics

When conducting RCTs, ethical considerations are important but rarely an obstacle. When considering RCTs to evaluate labour market policies, several key points emerge. First, public employment services regularly make complex decisions about assigning job seekers to programmes and interventions, aiming to provide them with the most suitable support. However, these decisions are often made in the face of significant uncertainty and without robust empirical evidence to guide them. Second, public employment services typically operate under binding resource constraints, meaning that not all job seekers can receive a given policy treatment.

Third, even though there is substantial research on active labour market policies, there are still many areas where evidence is lacking. For example, modern labour market policies may introduce new interventions that have not been previously evaluated. Evidence from one country may not necessarily generalise to other the Nordic countries. Policy designs may evolve over time, raising questions about their current effectiveness. Moreover, policymakers may wish to understand not only average effects but also heterogeneous effects across specific subgroups.

Taken together, these three factors provide strong justification for the use of RCTs. When tough allocation decisions must be made and resources are limited, random assignment can be as ethical as any other selection method. It also offers the additional benefit of generating credible causal evidence. This is particularly true when resource constraints mean that there are more job seekers deemed suitable for a given intervention than available slots. In other words, if there are

explicit or implicit queues for a programme, determining access through random assignment may be just as ethical as attempting to make difficult prioritisations among job seekers. This is especially relevant when it is unclear whether a given programme helps or harms job seekers' chances of finding employment. In such cases, it may even be unethical not to conduct an RCT, since without experimental evidence, public agencies may continue implementing programmes that actually harm job seekers.

However, RCTs may become ethically problematic if these conditions are not met. For example, if there is strong prior evidence that a policy has positive effects and sufficient resources exist to treat everyone, then it would be unethical to randomise and withhold treatment from some job seekers. Similarly, legal or policy rules may restrict randomisation. For instance, under Sweden's *Nystartsjobb* wage subsidy, employers are legally entitled to the subsidy if they hire a job seeker who has been unemployed for more than 12 months, which prevents randomisation.

Ultimately, these ethical considerations primarily concern the organisation implementing them (typically the public employment service), but researchers involved in RCTs must also carefully reflect on them. Formal ethical reviews should also be conducted if any sensitive personal data is processed.

3.3 Randomisation designs

A traditional RCT randomly assigns half of the units (e.g., job seekers) to a treatment group and the remaining half to a control group, which typically receives baseline services. Such RCTs estimate the effect of the evaluated treatment relative to the baseline, capturing the policy-relevant question of the additional direct impact of introducing the intervention. This standard design is a major improvement to non-experimental evaluations, but in some cases, it may also be valuable to consider more elaborate randomisation designs. One example is the displacement effects studied by Cheung et al. (2025), and the idea that early meetings that help some job seekers find jobs may displace jobs for other job seekers. Similar displacement effects may occur for other labour market policies, for instance, if training programmes or wage subsidies help treated job seekers secure jobs that would otherwise go to untreated individuals. To capture such displacement effects, randomisation at different levels, as in Cheung et al. (2025), may be important.

Two-level randomisation is one example of a more elaborate randomisation design. Other advanced designs include adaptive randomisation, which integrates random assignment with machine learning to improve allocation. Two-level randomisation can also be structured in various ways—not only across job seekers and local offices, but also across other relevant levels depending on the policy context. In cases with more than one type of treatment, such as in Bjørvatn et al. (2021), where a goal-setting intervention and a goal-setting intervention plus SMS

reminders are the two treatments, it may be optimal, in terms of statistical precision, to randomise more than half of the job seekers to the treatment.

3.4 Pre-analysis plans

When designing and evaluating RCTs, a pre-analysis plan (PAP) that describes the programme/intervention, the experimental design, research hypotheses, as well as empirical models and outcomes, is important for several reasons. This is particularly true for RCTs when the researchers are involved in the design and implementation of the intervention. One advantage of PAPs is that they require the researchers to clearly define hypotheses, outcome measures, and statistical methods before the experiment begins. By committing to these decisions in advance, PAPs reduce the risk of bias, selective reporting, and data mining. This strengthens the credibility, transparency, and replicability of experimental results, ensuring that findings reflect true effects rather than choices made after viewing the data.

In many fields, a pre-analysis plan is now also a requirement for publication in leading academic journals. However, some public policymakers may find PAPs restrictive, as they may wish to conduct exploratory analyses that address questions not specified in advance. While such flexibility is understandable, deviating from the PAP without documentation can undermine the credibility of the RCT, which is unfortunate since the central purpose of an RCT is to provide credible causal evidence on socially relevant questions.

3.5 Implementation challenges

Preparing and managing RCTs is often highly challenging, yet such challenges are rarely documented in research outputs. One exception is Cheung et al. (2020), which provides a dedicated implementation report for the RCT on early meetings evaluated by Cheung et al. (2025). This implementation report discusses some of the challenges.

First, as several other RCTs, Cheung et al. (2025) had an ambitious design intended to measure displacement effects, which required the mandatory participation of many local offices. Since participation was not voluntary in the first place, some offices lacked motivation, which made implementation more difficult. Opting for voluntary participation and randomising among local offices willing to take part in the experiment might have improved engagement and eased implementation. Similar issues can arise in other RCTs when the experimental protocol imposes specific requirements on caseworkers, local managers, or municipalities, potentially creating resistance or compliance issues. At the same time, only studying offices (or other types of participants) that enrol voluntarily may create issues related to external validity if the volunteers are very different compared to other offices.

Second, many RCTs struggle to recruit enough participants. To ensure adequate sample sizes, the randomisation procedure must be applied consistently to the entire target group. If caseworkers are responsible for randomisation as part of their regular duties, the process must be simple, well-integrated, and clearly communicated. Automating randomisation as much as possible can also help. Likewise, effective tools to monitor the implementation of RCTs are crucial. In this respect, Cheung et al. (2020) conclude that future large-scale RCTs should use integrated data monitoring systems with real-time feedback and supportive supervision to ensure that the RCTs are fully implemented in accordance with the design of the experiment.

Third, successful implementation requires strong organisational anchoring and engagement from both operational staff and local offices. Cheung et al. (2020) highlight that the local offices involved in the trial received resources to conduct more frequent meetings but sustained local management support proved critical. Everyone involved needed to understand the purpose and benefits of the project. In particular, time invested in explaining the value of randomisation paid off by increasing acceptance and compliance with the experimental design.

These lessons underscore the importance of anchoring the RCT across the entire public employment agency implementing it and establishing a dedicated implementation team with sufficient resources to monitor all stages of the trial. Beyond ensuring compliance and proper delivery of the intervention, embedding serves other essential purposes. First, designing an effective RCT often requires detailed institutional knowledge, which only insiders within the employment agency can provide. Second, if the intervention is not aligned with existing structures, the results may fail to generalise, or the trial may encounter unforeseen barriers. Third, when an RCT is well-integrated into the agency's operations, the lessons learned are more likely to lead to sustainable policy improvements. Fourth, evaluating and interpreting the results often depends on access to detailed data and a deep understanding of the institutional context.

3.6 Building institutions that support RCTs

The overview of RCTs in the Nordic countries shows that Sweden has conducted several RCTs over the past decade. This development is linked to the presence of a dedicated group within the Swedish Public Employment Service, whose staff have been strongly motivated to generate policy-relevant causal evidence through RCTs, as well as to their close collaboration with researchers at several academic institutions. This approach has proven highly successful. Most likely, without it, few, if any, RCTs on labour market policies in Sweden would have been implemented during the past decade (although the extent to which RCT activity would have differed in its absence is difficult to assess). The dedicated team within the Swedish PES, together with its collaboration with researchers, have contributed to motivating RCTs within the organisation, supporting careful implementation, and promoting rigorous scientific design and

continuous improvement of RCT-based evaluations. This has led to important evidence on policies such as early meetings with caseworkers (Cheung et al., 2025), private providers (Egebark et al., 2024), online job recommendations (Hensvik et al., 2025), supported employment (Fogelgren et al., 2023), early intervention programmes and job-search assistance initiatives for immigrants (Dahlberg et al., 2024; Helgesson et al., 2020), as well as light-touch interventions using email communication (Cockx et al., 2025).

The Swedish experience suggests that it may be valuable to build institutions that systematically support RCTs in the Nordic countries. Rather than relying primarily on collaboration between a few highly motivated individuals, more durable institutional frameworks could facilitate more rigorous experimentation and learning about the causal effects of labour market policies. At the same time, several challenges are likely to complicate such developments. For example, some actors may be hesitant to rely on evidence from RCTs, especially when such evidence calls into question ideologically motivated policies or long-standing practices within the Public Employment Service. Some actors may also feel so confident in a policy's effectiveness that additional causal evidence is regarded as unnecessary. In addition, as discussed above, implementing RCTs can be resource-intensive and require careful preparation and trial management.

Motivating and sustaining institutions capable of addressing these obstacles is itself challenging. Although Nordic contexts differ, building such institutions may benefit from collaboration between ministries, public employment agencies, and academic institutions. Explicit support for RCTs within these organizations, combined with joint efforts to establish more durable institutional frameworks, could help facilitate experimentation and learning beyond the involvement of a small number of dedicated individuals.

The Swedish experience also shows that close collaboration between public employment agencies and academic researchers can be an effective way to implement and learn from RCTs. Owing to this collaboration and the close involvement of researchers within the Swedish Public Employment Service at every stage of the RCT process, several evaluations have been co-authored by researchers from the Swedish PES and academic institutions (Cheung et al., 2025 is one example). This has received some criticism about the independence and credibility of the research output. However, several safeguards can mitigate such concerns: 1) A detailed pre-analysis plan, submitted before the analysis, ensures transparency and prevents undue influence being exerted on the evaluation; 2) Formal contracts between independent researchers and public employment agencies should specify the obligations of all parties and give the researchers the right to complete the evaluation autonomously if the public employment agency interferes with the evaluation. Moreover, implementing an RCT inherently requires significant involvement from

public employment agencies, and any manipulation by them may occur during the implementation phase. Thus, what matters most for credibility is probably not excluding public agency staff from co-authorship, but ensuring transparency, independent analysis, and clear safeguards throughout the research process.

4 Analysing RCTs of labour market policies

Besides solving selection problems and providing credible evidence, RCTs also tend to simplify analysis, as they typically involve comparing the average outcomes of the randomised treatment and control groups. However, by boosting take-up, using objective outcome and cost data, exploiting high-quality Nordic administrative data, and carefully considering general equilibrium effects, additional policy lessons can be learned from RCTs. These aspects are discussed in this section.

4.1 Take up

In a standard RCT, take up refers to the difference in actual treatment participation between the randomised treatment and control groups. It depends on the share of individuals in the treatment group who ultimately receive the intervention and the share in the control group who receive it despite not being randomised to treatment. As such, it is closely related to the above-mentioned implementation challenges, and a high take-up rate usually indicates that the RCT was implemented according to the pre-analysis plan. At the same time, it is important to note that take-up rates well below 100% are quite common, but in most cases, this does not prevent meaningful analysis.

The RCTs mentioned in Section 2 all delivered robust and credible evidence on many policy-relevant questions, despite substantial differences in take up. For the early meetings experiment in Cheung et al. (2025), the take-up rate was 23%, meaning that 23% of the job seekers in the randomised treatment group attended extra meetings. This can be compared with a take-up rate of 100% in the light-touch habits intervention in Bjørvatn et al. (2021), as the analysis sample was defined as those who completed the online survey and the intervention. For the private-provider intervention in Rehwald et al. (2017), take up is slightly more complex, as the experiment ended prematurely in June 2012. This meant that the entire control group—who should have received public services but were eventually transferred to private providers—were also treated toward the end of the experiment. However, this does not invalidate the design, as the RCT could still measure the impact of private versus public support up to the end of June 2012. During this time, the difference in treatment exposure remained sizeable.

The take-up rates also vary across other Nordic RCTs, but this does not prevent meaningful analysis as long as there is a sufficiently large difference in take up between the treatment and control groups and if the sample size is large enough to detect a meaningful effect. For instance, a take-up rate of 20–30% can still yield valid results in a large sample. In such cases, there remains a clear difference in treatment exposure, and any observed differences in outcomes between the treatment and control groups can still be credibly attributed to the intervention. In such cases, it is, however, instructive to also document and describe the various reasons behind any take up that is less than 100%. It helps to understand the implementation of the RCT and provides information about the average effect captured in the evaluation.

In these cases, traditional questions about which treatment effect to estimate become especially relevant. One option is to estimate the intention-to-treat (ITT) effect—the “reduced-form” impact of being assigned to treatment—measured as the simple difference in outcomes between the treatment and control groups. Since, most policies are not mandatory this may also be the most policy relevant effect to study.

At the same time, it is important to point out that high take-up is highly desirable, as it increases the statistical power of the experiment. When more job seekers in the treatment group actually receive the intervention, the evaluation captures more information about treatment outcomes, reducing uncertainty around the estimated average effects. By contrast, if there is no significant difference in take up between the treatment and control groups, the RCT has effectively failed, as there is no real variation in treatment exposure. In such cases, any observed differences in outcomes cannot be attributed to the policy being evaluated.

4.2 Objective outcomes

RCTs are designed to provide credible evidence on the effects of a policy. To accomplish this, it is crucial to use objective outcome measures that do not contaminate the RCT. The choice of outcomes depends on the data sources available, which may include self-reported assessments, caseworker reports, survey data, administrative records, or external sources such as tax data. Each type has strengths and weaknesses, but several general insights apply.

First, administrative data on unemployment spells collected by public employment agencies can pose challenges. These registers typically rely on caseworkers and job seekers accurately reporting when spells of unemployment end and why. Problems arise when incentives to report accurately are influenced by the intervention itself. For example, in evaluations of private providers, reporting responsibilities may shift to the providers, who could have incentives to record shorter unemployment spells if performance-based payments are tied to such outcomes. Indeed, in a recent evaluation of private providers in Sweden, Egebark et al. (2024) finds substantial differences between in-house unemployment data and more objective tax records that

cannot easily be manipulated. However, when the treatment does not affect reporting incentives, unemployment data from public agencies are more reliable.

Second, survey data can complement administrative data by capturing outcomes not observed in registers, such as job quality, health responses, or mechanisms underlying observed effects. Yet surveys face increasing challenges due to low response rates, which reduces the effective sample size—for example, if only 10% respond—and may introduce bias if response likelihood differs by treatment status, undermining the RCT. Thus, while adding new data, surveys pose a challenge when used to construct key outcome data. Note also that in some settings, it is possible to achieve high response rates. One example is the habits intervention in Bjørvatn et al. (2021), which conducted follow-up surveys one and six months after the intervention and obtained response rates of 63% and 66%, respectively. One factor that probably contributed to the high response rates is that the intervention sample consisted of job seekers who had completed an initial online survey. Nevertheless, the high response rate is still impressive.

Third, some outcomes are conditional, meaning they are only observed for a subset of participants. In labour market studies, a common example is the wage of the first job after unemployment, which is only available for individuals who found jobs. If job-finding rates differ between the treatment and the control group, comparisons of wages may suffer from selection bias. This is a well-known issue with no simple solution, but it highlights the importance of carefully selecting and interpreting outcome measures.

4.3 Costs

Measuring outcome differences—such as the gap in employment rates between treated and control groups—is, of course, important. However, it is also essential to consider the costs of an intervention. A small positive employment effect can lead to very different conclusions depending on whether the policy is cheap or expensive. Likewise, a negative employment effect may still be acceptable if the intervention is less costly than the baseline programme received by the control group.

That said, measuring costs is challenging, and several approaches can be used. From a strict cost–benefit perspective, all relevant costs (e.g., financial expenses, time, resources, or risks) and benefits (e.g., efficiency gains, improved outcomes, or social value) must be identified, assigned measurable—often monetary—values, and compared. In this framework, costs are all activities that consume resources, while benefits are those that create new resources. Thus, policies such as early meetings, training programmes, or rehabilitation measures count as costs, since they require staff time and related inputs. Conversely, wage subsidy programmes are not considered costs in the strict sense, as they merely reallocate resources from public budgets to private firms or individuals (abstracting from administrative and supervisory expenses). In other cases, whether

an expense is a cost depends on its economic nature. For example, reimbursements to private providers are costs when they finance services delivered to job seekers, but not when they generate profits, since strictly speaking, this represents a transfer from public to private actors rather than a net resource use. Similar complexities arise in measuring benefits. What is the economic or social value of reducing average unemployment duration by one month? The answer depends, among other things, on the production value associated with less time spent unemployed, which is sometimes difficult to quantify.

For these and other reasons, it is usually easier and in practice more common to measure impacts on public spending instead of a full-blown cost-benefit analysis. From such a public-spending perspective, the benefits of an intervention may come in the form of lower public expenditures (e.g., lower unemployment benefit payments) and public revenues (e.g., tax revenues). The costs are any increased public spending due to the policy of interest, including caseworker costs for early meetings, reimbursements to private providers, employment subsidies, etc. This public spending perspective differs from a strict cost-benefit analysis, as it focuses on impacts on budget spending rather than overall social welfare. However, a government spending analysis is still often policy-relevant, as the budgetary consequences of a policy are frequently of key importance to policy decisions. On a related note, often the focus is on budget spending (i.e., programme costs and unemployment insurance benefits), while ignoring budget revenue (e.g., tax revenues and social contributions).

In any case, when conducting RCTs, it is often important to take costs into account and to be explicit about the type of cost analysis (cost-benefit or public spending) being carried out. As already demonstrated, this distinction may have important implications for policy conclusions. To illustrate, consider the cost perspective in the three RCTs from Section 2. The habits intervention in Bjørvatn et al. (2021) was digital and self-guided, and thus almost costless to deliver once it had been designed. If scaled up, the cost per participant would be virtually zero, meaning that even small positive employment effects would be favourable from a cost-benefit perspective. However, there may still be limitations on the types and numbers of light-touch interventions job seekers can participate in.

Rehwald et al. (2017) conduct a detailed cost analysis focusing on the budgetary burden of private and public provision of employment services. The costs for the public programme can be divided into three components: 1) meeting costs in terms of caseworkers' time; 2) labour market programme costs; and 3) transfer payments. To quantify meeting costs, they follow a previous study that estimated the time spent on each meeting and the hourly wage cost of caseworkers. Programme costs are calculated using the average cost per participant-week. Transfer payments include unemployment insurance benefits, social assistance benefits, and other individual-level

benefits, as well as wage subsidies paid to firms. The costs of the private services follow a similar structure. They include direct payments made to private contractors, programme costs not borne by the private providers, and transfers, as in the public option. As already mentioned, the results of this cost analysis show that the private option is significantly more expensive while not producing better employment outcomes (partly because the analysis includes unemployment insurance and other benefits).

Finally, to evaluate the early-meetings program, Cheung et al. (2025) conduct a government spending analysis comparing reduced UI payments with programme costs. Benefit savings are estimated by multiplying the share of unemployed by their UI benefit levels under scenarios with and without the programme. Programme costs are the expenses for the caseworkers who conduct the early meetings. However, calculating these costs is challenging because the additional meeting expenses are not easily separable from regular caseworker costs at local offices; the accounting data only reports total caseworker costs. To overcome this, Cheung et al. (2025) use a time-use survey in which caseworkers reported how much time they spent on each meeting, including preparation, the meeting itself, and follow-up tasks such as documentation. Based on an average of 73 minutes of caseworker time per meeting, combined with information on caseworkers' monthly working hours, wages, and overhead, the authors calculate the cost per meeting and the total cost per participant for the early-meetings programme. This calculation demonstrates that it may be possible to obtain information about the costs of an intervention, even if the public employment agency does not explicitly separate programme costs from other costs. In terms of results, Cheung et al. (2025) find favourable government spending effects as the gains of the early meetings in terms of lower UI benefits outweigh the programme costs.

To summarise, the cost side is often key for policy but poses several challenges, for instance, related to separating the costs of a certain intervention from other costs incurred by public employment agencies. Hopefully, the examples above show that it is possible to conduct valuable cost analyses and that this should be an integral part of RCT evaluations, but this requires careful planning to assure that cost data is available for the evaluation.

4.4 Nordic data and analyses of mechanisms

Learning about employment effects and costs in a credible way is a major advantage of RCTs. In addition, several Nordic RCTs show that even more can be learned when these experiments are combined with detailed Nordic administrative data. The Nordic countries have a distinct advantage in terms of individual-level data, which can further enhance what we learn from RCTs. For instance, using detailed data can help uncover the mechanisms behind observed effects—or the lack thereof—that is, help us understand why a certain policy worked or failed. This, in turn, can improve the design of future interventions and make it easier to translate lessons learned from

RCTs to other contexts. Detailed Nordic data also enables the analysis of various additional outcomes, providing insights into relevant side effects of interventions, such as health consequences, the impact on families, and labour market exits, in addition to employment effects.

The researchers can achieve this in various ways. Take the early meetings evaluated by Cheung et al. (2025) as an example. They find that the early meetings boost job finding among the participants but also create displacement effects among non-participants in the same local market. The study also exploits the rich administrative data to study the mechanisms behind these effects. They exploit detailed administrative data from the Swedish PES on caseworkers (e.g., vacancy referrals and programme assignments) and job seekers (e.g., search behaviour). One result of this exercise is that early meetings are effective because they increase the number of vacancy referrals passed on from caseworkers to job seekers. Caseworkers use their expertise to find and point job seekers to relevant job openings. The jobseekers do not broaden their search in terms of occupations or geographical distance. Instead, the vacancy referrals lead to a more streamlined job search process by helping job seekers apply to the most relevant jobs earlier. The authors also find that face-to-face and distance meetings have equally positive employment effects, supporting the case for communication technologies when providing job search assistance.

This analysis uncovers the mechanisms behind the direct effects of early meetings for participants, but Cheung et al. (2025) also examine the mechanisms underlying displacement effects. Again, exploiting detailed Swedish data, they examine resource allocation and document that the negative effects for the non-treated do not arise because resources are allocated away from them—indicating that job displacement is the key mechanism. This finding is actually bad news for policy, as resource allocation issues can be mitigated by adjusting funding arrangements, whereas job displacement is far more difficult to address. In essence, the non-treated do not receive fewer referrals; rather, each vacancy receives more referrals, thereby increasing competition for available jobs and creating the observed displacement effects.

In summary, Cheung et al. (2025) use Swedish and Norwegian data, but similarly detailed and interesting administrative data are available in other Nordic countries. This is a great strength of labour research in the Nordic countries, a strength that is important to exploit when evaluating RCTs.

4.5 General equilibrium effects

The discussion of Cheung et al. (2025) and early meetings in Section 2.2. highlights the importance of general equilibrium (GE) effects. It includes job displacement, effects on wages and vacancies, and other more general responses to a given policy. Therefore, it is important to think about GE effects when designing RCTs. Key questions include: Is job displacement likely to be significant for the policy being studied? Could the intervention affect wages or the creation

of vacancies? Might it affect how job seekers apply for jobs, and might such changes spill over to other job seekers? Several studies from outside the Nordic countries have also shown that accounting for such GE effects may be highly important for assessing the cost-effectiveness of labour market policies (Crépon et al., 2013; Gautier et al., 2018; Ferracci et al., 2014).

In addition to the more specific discussion of the general equilibrium effects of early meetings and other job-search assistance interventions in Section 2.2, several general points are worth noting. First, in some settings, GE effects may be extremely important to consider, as they can lead to very different policy conclusions, while in other settings, GE effects may be of second-order importance. Second, GE effects may either reinforce direct positive employment effects (e.g., the creation of new vacancies) or mitigate them (e.g., through job displacement). Third, in some contexts, it may be possible to capture GE effects within the randomisation design (e.g., a two-level design). Fourth, analysing GE responses in terms of vacancies and wages and incorporating them into structural analyses may enable more elaborate policy simulations and deeper policy learning beyond the original RCT.

5 Concluding remarks

This paper has used selected examples of randomised controlled trials from the Nordic countries to highlight central issues in motivating, designing, implementing, and evaluating various labour market policies using RCTs. Rather than aiming to provide an exhaustive review of the literature, the paper has illustrated how RCTs can be applied in practice and what can be learned from their use in real-world policy settings. One conclusion is that RCTs offer a powerful and often feasible approach for evaluating labour market policies. When resources are limited, and allocation decisions need to be made, random assignments can be ethically defensible and, in many cases, preferable to alternative allocation mechanisms. At the same time, the discussion in this paper emphasises that the credibility and usefulness of RCTs hinge on careful design and implementation, including transparent randomisation procedures, attention to ethical concerns, research independence, and practical constraints faced by the agencies implementing them.

Going forward, it is important to realise that conducting and evaluating RCTs can be challenging, and establishing institutions and collaborations that foster more RCTs and enable effective learning from them is important for all of the Nordic countries aiming to improve the efficiency of their labour market policies. Continued collaboration between researchers and policy institutions, combined with careful attention to design, implementation, and interpretation, will be essential for fully realising the potential of RCTs to inform effective labour market policy.

The paper also underscores that conducting and analysing RCTs involves more than simply comparing outcomes between treatment and control groups. Issues such as take-up, compliance,

outcome measurement, and cost-effectiveness are central to interpreting results and informing policy decisions. The Nordic countries' rich administrative register data constitute a major advantage in this respect, enabling precise measurement of outcomes over time and across populations. However, the presence of general equilibrium effects, such as displacement of jobs, complicates interpretation, as positive effects for treated individuals may be partly offset by displacement or spillovers affecting non-treated job seekers.

Overall, the paper highlights both the promise and the limitations of RCTs in labour market policy evaluation. On the one hand, RCTs are generally regarded as the gold standard of evaluation methods, and in some cases, methods based on non-experimental data may lead to incorrect policy conclusions about the effectiveness of key policies. One example is research on private providers, where some non-experimental approaches suggest positive effects, whereas carefully conducted RCTs have shown that private providers do not appear to be a cost-effective way to deliver labour market policies. This illustrates that RCTs are a crucial tool for generating credible evidence on what works, for whom, and under what conditions.

On the other hand, it is important that evidence from RCTs is complemented by evidence from other methodological approaches. In some cases, it is possible to exploit natural experiments to estimate causal effects in a credible way. For instance, if a policy is implemented only in certain areas, this may allow for credible comparisons across local areas, potentially also exploiting data from before the policy was introduced. In addition, several studies have exploited caseworker variation to obtain as-if random variation in policy assignment, effectively replicating an RCT without explicit randomisation (Cederlöf et al., 2025; Humlum et al., 2025). There may also be situations in which RCTs are not feasible, for example, due to ethical considerations, resource constraints, or institutional regulations. In such cases, non-experimental methods are often preferable to not evaluating.

References

- Altmann, S., Falk, A., Jäger, S., Zimmermann, F., (2018), “Learning about job search: a field experiment with job seekers in Germany”, *Journal of Public Economics*, 164, 33–49.
- Berg, H., K. E. Hauge, S. Markussen and T. Zhang (2021), “Supported Employment eller vanlig oppfølging? Resultater fra et stort randomisert forsøk i NAV”, Rapport 2021:2. Oslo: Frischsenteret.
- Benmarker, H., Grönqvist, E. and Öckert, B. (2013), “Effects of contracting out employment services: evidence from a randomized experiment”, *Journal of Public Economics*, 98, 68–84.
- Belot, M, P Kircher and P Muller, (2019), “Providing advice to jobseekers at low cost: An experimental study on online advice”, *Review of Economic Studies*, 86:1411–47.
- Bjorvatn K, M Ekström and A J. Garcia Pires (2021), “Setting goals for keystone habits improves labor market prospects and life satisfaction for unemployed youth: Experimental evidence from Norway”, *Journal of Economic Behavior & Organization*, 188, 1109-1123.
- Blasco S and M Rosholm (2011), “The impact of active labour market policy on post-unemployment outcomes: Evidence from a social experiment in Denmark”, mimeo.
- Cairo S and R Mahlstedt (2023), “The disparate effects of information provision: A field experiment on the work incentives of social welfare”, *Journal of Public Economics*, vol 226.
- Card, D, J Kluge and A Weber (2010), “Active labour market policy evaluations: A meta-analysis”, *Economic Journal*, 120: F452–F477.
- Card, D, J Kluge and A Weber (2018), “What works? A meta analysis of recent active labor market program evaluations”, *Journal of the European Economic Association*, 15(3), 894–931.
- Cederlöf J, M Söderström and J Vikström (2025), “The Role of Caseworkers: Job Finding, Job Quality and Determinants of Value-Added”, *Journal of the European Economic Association*, accepted.

Cheung M, J Egebark, A Forslund, L Laun, M Rödin and J Vikström (2020), “Implementation of a labor market program with more frequent meetings in Sweden”, IFAU Working Paper 2020:22.

Cheung M, J Egebark, A Forslund, L Laun, M Rödin and J Vikström (2025), ”Does Job Search Assistance Reduce Unemployment? Evidence on Displacement Effects and Mechanisms”, *Journal of Labor Economics*, 43:1, 47-81.

Crépon, B, E Duflo, M Gurgand, R Rathelot, and P Zamora (2013), “Do labor market policies have displacement effects? Evidence from a clustered randomized experiment”, *Quarterly Journal of Economics*, 128(2), 531–80.

Cockx B, J Egebark, G Van Hoyer, E Videnord, J Vikström (2025), “Motivating Job Seekers. A Field Experiment”, mimeo IFAU.

Dahlberg M, J Egebark, U Vikman and G Özcan (2024), “Labor market integration of refugees: RCT evidence from an early intervention program in Sweden”, *Journal of Economic Behavior & Organization*, 217, 614-630.

Engström, P., Hesselius, P. and Holmlund, B. (2012), Vacancy Referrals, Job Search, and the Duration of Unemployment: A Randomized Experiment, *LABOUR*, 26: 419-435.

Egebark J, L Laun, L Liljeberg, M Rödin, M Söderström E Videnord and J Vikström. 2024. En effektutvärdering av arbetsförmedling med fristående leverantörer (An impact evaluation of employment services with independent providers), IFAU Rapport 2024:17.

Ferracci, M, G Jolivet, and G J. van den Berg, (2014), ”Evidence of treatment spillovers within markets”, *Review of Economics and Statistics*, 96(5), 812–23.

Fogelgren, M, P Ornstein, M Rödin and P Skogman Thoursie (2023), ”Is Supported Employment Effective for Young Adults with Disability Pension? Evidence from a Swedish Randomized Evaluation”, *Journal of Human Resources*, 58, 452–487.

Gautier P, P Muller, B van der Klauuw, M Rosholm and M Svarer (2018), “Estimating equilibrium effects of job search assistance”, *Journal of Labor Economics*, 36(4):1073–1125.

Graversen, B. and van Ours, J. (2008a), “Activating unemployed workers works: experimental evidence from Denmark”, *Economic Letters*, 100, 308–310.

Graversen, B. and van Ours, J. (2008b), “How to help unemployed find jobs quickly: Experimental evidence from a mandatory activation programme”, *Journal of Public Economics* 92, 2020–2035.

Helgesson P, E Jönsson, P Ornstein, M Rödin and U Westin (2020), “Equal Entry – can job search assistance increase employment for newly arrived immigrant women?”, *Arbetsförmedlingen analys* 2020:10.

Hensvik L, T Le Barbanchon and R Rathelot (2025), “Recommender systems and the labor market”, mimeo Uppsala University.

Hernæs O (2025), “Performance pay for private program providers and impact on participants: A field experiment with employment services in Norway”, *Labour Economics*, vol 95.

Høeberg L, Poulsen J, Hertz M, Svarer M & M Rosholm (2011), "Evaluering unge - godt i gand", Report Ramboll.

Humlum, A., J.R. Munch, and M. Rasmussen (2023). “What Works for the Unemployed? Evidence from Quasi-Random Caseworker Assignments.” IZA discussion paper no. 16033.

Hägglund P. (2014) “Experimental evidence from active placement efforts among unemployed in Sweden”, *Evaluation Review*, 38(3), 191-216.

Hägglund, P. (2011), Are There Pre-Programme Effects of Swedish Active Labour Market Policies? Evidence from Three Randomized Experiments, *Economic Letters*, 112, 91–93.

Joona, P. A. and L Nekby (2012), “Intensive Coaching of New Immigrants: An Evaluation Based on Random Program Assignment.” *The Scandinavian Journal of Economics*, 114(2).

Laun, L. and Thoursie, P. (2014), “Does privatisation of vocational rehabilitation improve labour market opportunities? evidence from a field experiment in Sweden”, *Journal of Health Economics* 34, 59–72.

Maibom, J., Rosholm, M. and Svarer, M. 2017, Experimental Evidence on the Effects of Early Meetings and Activation“, *Scandinavian Journal of Economics*, 119: 541-570.

Malmberg-Heimonen, I. and Vuori, J. (2005), “Financial Incentives and Job-search Training: Methods to Increase Labour Market Integration in Contemporary Welfare States?”, *Social Policy & Administration*, 39: 247-259.

Pesola, H, Sarvimäki, M and Virkola T (2025), “Randomization as an Incentive Device: Evidence from Public Procurement of Immigrant Integration Services”, available at SSRN.

Rehwald K, M Rosholm and M Svarer (2017), “Do public or private providers of employment services matter for employment? Evidence from a randomized experiment”, *Labour Economics* 45 (2017) 169–187.

Rosholm M. (2008), “Experimental evidence on the nature of the Danish employment miracle”, Working Paper 08-14. University of Aarhus.

Rosholm M and M. Svarer (2009a), "Kvantitativ evaluering af Hurtig i gang 2", Report Arbejdsmarkedstyrelsen.

Rosholm M and M. Svarer (2009b), "Kvantitativ evaluering af Alle i gang", STAR Arbejdspapiret.

Sørensen, K. L. (2016), “Heterogeneous impacts on earnings from an early effort in labor market programs”, *Labour Economics*, 41, 266-279.

Sørensen, K. L. (2017), “Active labour market programmes and reservation wages: it is a hazard”, *Applied Economics Letters*, 24(9), 589–593.

SOU (1978), “Arbetsmarknadspolitik i förändring studier kring arbetsmarknadspolitikens verkningar: betänkande av expertgruppen för utredningsverksamhet i arbetsmarknadsfrågor (EFA)”, SOU 1978:60.

Sveinsdottir, V., T. Fyhn, M. Frangakis and J. Opsahl (2020). "Effektevaluering av Raskt i jobb for flyktninger - En randomisert kontrollert studie", Rapport 2020:3. NORCE Norwegian Research Centre.

Vinokur, A. D., Schul, Y., Vuori, J., and 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(1), 32.

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(1), 5.

Vikström J, Michael Rosholm, Michael Svarer (2013), "The effectiveness of active labor market policies: Evidence from a social experiment using non-parametric bounds, *Labour Economics*, 24,58-67.