Chapter 16
Initiatives to Combat the Labour Market Exclusion of Youth in Northern Europe: A Meta-analysis

Inés Hardoy, Knut Røed, Kristine von Simson, and Tao Zhang

16.1 Introduction

The high youth unemployment in Europe in recent years has called both for the implementation of more measures to enhance employability and a pressing need to better understand their impact. There is an extensive literature that studies the effects of active labour market programmes, for the population at large and for certain target groups, in particular, not only in Europe but also in the USA and in the rest of the world. However, it is difficult to synthesize, interpret and draw conclusions directly from the literature available because studies vary along many different dimensions: with respect to the type of programme, the target group, the economic conditions, the method of evaluation, etc.

A tool that can be used to compile and compare results from several empirical studies is so-called meta-analyses. It entails putting together evaluation studies in a synthetic database, establishing a common measure of performance, identifying factors associated with positive and negative impacts and then using standard regression techniques to analyse the various factors individually contributing to the estimated results. This enables the researcher to get clearer, more robust and more statistically significant results than can be obtained from each individual study.[1]

Meta-analyses have been widely used in medical and biological studies, where data is generated in controlled experiments/trials, and each single study often consists of a limited number of observations, diminishing the empirical predictive power of the estimates. Meta-analyses have also, gradually, been more used to

[1] For an introduction to meta-analysis, see, e.g. Stanley and Doucouliagos (2012).

I. Hardoy (✉) • K. von Simson
Institute for Social Research, Oslo, Norway
e-mail: ines.hardoy@samfunnsforskning.no

K. Røed • T. Zhang
Frisch Centre, Oslo, Norway

M.Á. Malo, A. Moreno Mínguez (eds.), *European Youth Labour Markets*,
https://doi.org/10.1007/978-3-319-68222-8_16

collect information across studies in the social sciences as well. More recently meta-analyses have been used to study the impact of labour market programmes as well.

An important conclusion in a meta-analysis covering more than 200 studies (Card et al. 2015) is that programmes that create incentives in the private sector, such as wage subsidies and assistance in job search, are usually associated with favourable effects. The same holds for training (both classroom courses and job training), while placement in the public sector has on average, a negative effect on employment. Both Card et al. (2015) and Kluve (2010) cover the population at large and programmes implemented both in Europe and in other countries. Both articles study the impact for young people in particular as well, by including youth programmes as a regressor, and conclude that youth is a particularly difficult group to assist. The meta-analysis of Greenberg et al. (2003), based on the US data covering the period until the end of the 1990s, arrives at similar conclusions. Puerto (2007) covers studies of youth programmes from around the world and also concludes that there are little or no positive effects of labour market programmes for youth. In contrast, a very recent study focusing exclusively on youth programmes is more optimistic. The meta-analysis by Kluve et al. (2016) systematizes 113 evaluations of youth measures from around the world and concludes that about one-third of the measures evaluated have statistically significant positive effects. However, the effects appear to be most positive for measures implemented in countries with medium and low income levels.[2]

Compared to the meta-analyses mentioned above, our aim is to go more thoroughly into a group of countries that are homogeneous in many respects and also draw from a larger selection of studies, both unpublished and published in referee journals, written in English or in the native language of the country. The countries included are Finland, Sweden, Norway, Denmark, Germany and the UK. We focus on initiatives/programmes that seek to improve the labour market prospects of unemployed youth below age 30. We aim to answer the following questions:

- How do the estimated effects vary with programme type and over time?
- How do impacts vary with programme-related characteristics, e.g. differences in target group, and with methods of evaluation?
- How do contextual factors, such as macroeconomic conditions and institutional differences affect the estimated effects?

The paper proceeds as follows: In the next chapter, we describe the data compiled and the methodological approach. Thereafter we present descriptive statistics followed by the results. We conclude with a summary and some closing remarks.

[2] The conclusions of Kluve et al. (2016) are consistent with a recent comprehensive study of youth measures in Europe conducted by Caliendo and Schmidl (2016), but note it is not a meta-analysis.

16.2 Data and Methodological Approach

The studies included in our database have undergone a careful scrutiny. We include only studies using reduced-form methods attempting to identify causal effects of interventions, taking into account selectivity in recruitment.[3] We take the results of the individual studies at face value. We do not require that the paper is written in English, i.e. we include studies written in the native language. We use different web search engines, including bibliographic databases such as IDEAS and Google search, as well as working paper series of institutions such as IZA, IFS, IFAU and NBER that publish high-quality working papers before their actual publication in a journal. The reason for the more liberal approach is an attempt to avoid potential publication bias in the sense that analyses showing effect are more likely to be published than those non-significant effects. To be included in our database, they must have been published from 1999 onwards, and we include only evaluations of measures implemented from the beginning of the 1990s to 2013. We do not require the studies to be published in refereed journals. The studies can be reports, working papers, own memo from a research institute, the EU/Nordic countries ministerial council publications, etc. In total there are 44 studies from the six countries included in our database. A full list of all studies is included in the reference list.

It is important to find a common measure and statistical method to analyse the estimated effects of programmes across studies and countries, over time. We follow the methods in Card et al. (2010), Kluve (2010) and Card et al. (2015), using a so-called "ordered probit" (OP) regression. The dependent variable in the meta-analysis is the estimated outcome effects of programmes. Now, different studies measure the success/failure of programmes differently, or they evaluate various outcomes of one and the same programme. Some focus on job probabilities while others measure the impact of an intervention on wage income. Some look at the impact on reduced welfare dependence, while others look at transitions to ordinary education. Different outcomes thus provide measures of different effects. To account for this diversity, we define a variable *outcome* to indicate which outcome measure is used to evaluate the impact of the programme. Possible outcomes are *employment*, *unemployment*, *wage income*, *welfare dependence* and *education*.

Since it is not possible to compare the sizes of the estimated effects of the different studies directly, we follow Kluve (2010) and define a latent outcome effect variable Y^*, which is a continuous normally distributed variable. We define a categorical variable Y which takes three discrete values: $Y = 1$ if the outcome effect is significantly negative, $Y = 2$ if the outcome effect is not statistically significant and $Y = 3$ if the effect is significantly positive. If a programme increases the probability of obtaining a job, it is regarded as having a positive effect. Likewise, if it reduces the likelihood of welfare dependency or unemployment, then the programme is also

[3] One key issue in evaluation studies is that the counterfactual, i.e. what would have happened in the absence of an intervention, is not observed. Had we had that piece of information, we could have just taken the difference and inferred that that is the impact.

regarded as having a positive effect.[4] OP is then used to estimate the effects of explanatory variables on the probability distribution function of Y^*. From these estimates we can derive the marginal effects of each of the explanatory variable on the probability of positive/negative programme outcomes.

We divide active labour market programmes targeting youth in six categories:

1. *Training* entails classroom courses and other off-the-job training programmes organized by labour offices or private agents.
2. *Work practice* is a group of programmes providing work experience and practical know-how and includes training provided while at work (on-the-job training).
3. *Wage subsidies*, primarily in the private sector.
4. *Public employment* measures that give practice/job opportunity in the public sector (in the case of youth with reduced work capacity in sheltered enterprises).
5. *Intensified activation* includes monitoring of search activity, mentoring/supervision and close monitoring but also the use of "threats" and sanctions. This category also includes the "youth guarantee", since this initiative is intended to provide early efforts to fight youth unemployment and preventing marginalization and reduce the proportion of youth "not in employment, education or training" (NEET)
6. *Other programmes* is a residual category comprising "outsourcing of employment services" to private providers ("decentralizing measures") and start-up subsidies for self-employment.

We distinguish between short-term and long-term effects. If the impact is evaluated within the first 12 months of completion, we define it as a *short-term effect*. When the evaluation covers a period beyond 12 months after the programme is completed, we define the effect as *long-term*.

We define a variable *method* to classify the method used in the evaluation. The method considered to be the gold standard for causal inference is *randomized control trials*. Experiments are conducted such that individuals are randomly assigned to a treatment and control group, and provided that everything works by the book, taking differences of outcomes between treated and controls give the causal impact of the treatment. Another category, labelled *diff-in-diffs*, includes studies that primarily rely on quasi-experimental methods such as difference-in-difference and regression discontinuity design, as well as instrumental variable methods and two-stage least square. These methods are used when an exogenous event occurs, such as a reform or intervention, which randomly splits the group of interest such that part of the group is affected by the intervention while the other is not. *Matching* involves establishing a comparison group of non-participants which resembles as much as possible an already established participant group, so that differences in outcomes can be attributed to the measure. In this category we also include studies which use a combination of matching and difference-in-differences. We also include

[4] Note that we use the terms "positive" and "negative" in a normative – and not a mathematical – sense throughout this paper.

a category for duration methods or *timing-of-event analysis* where what is modelled is the duration until a certain change, such as a transition from unemployment to employment occurs. Finally, we have a residual category with *other methods* including linear regression and structural models.

As mentioned above we include evaluations of measures implemented from the beginning of the 1990s to 2013, i.e. a relatively long period of time. Originally we wished to divide studies by periods. However, it turns out that many studies cover more than a 10-year period, and many studies overlap between periods making the creation of several intervals inappropriate. Hence we create a dummy variable that describes whether the period of programme implementation is *before* or *after 2004*. The year 2004 is chosen arbitrarily; it divides the period of analysis roughly in half.

Also there are some studies that evaluate measures for different groups of young people, for example, by evaluating effects separately for men and women, for those with health disadvantages or for those with unemployment benefits and/or on social assistance. We define dummy variables to indicate that evaluation is done for specific subgroups. This enables us to investigate whether focusing on specific subgroups gives rise to more positive or negative effects than analysing the entire population.

Several studies point to the close link between labour market institutions, such as hiring/redundancy policies and minimum wages, and youth unemployment (e.g. Jimeno and Rodrigues-Palenzuela 2003; Bertola et al. 2007). To capture macroeconomic conditions, we calculate the country-specific rate of *youth unemployment* (in percentage) averaged over the period that the programme being evaluated was implemented. In addition, we use three variables to characterize some important institutional differences between countries that can potentially affect both school-to-work transitions and youth employment conditions: (i) indices of the *strictness of employment protection* at the individual level, (ii) rigorousness of temporary employment regulation and (iii) total public expenditure on active labour market policies (ALMP) as percentage of GDP. The institutional variables are also measured at the country level as an average over the period the programme being evaluated was implemented.[5]

We define an observation (a data point) in the meta-database as a unique combination of programme type, evaluation method, country, if there is a short-term/long-term effect, etc. Each study may contribute with additional data points if several methods are applied, or different outcomes are investigated, or different programmes are evaluated, etc. Hence, the number of observations in our database is a lot larger than the number of studies in the database. Since observations from one and the same study cannot be perceived as totally independent from each other, we cluster by study.

Table 16.1 shows descriptive statistics for each of the countries and for all effect estimates taken together. As the last column shows, there are 44 studies and 425 observations in our database. Germany has the largest number of studies, followed by Sweden. Denmark has the largest share of observations indicating positive

[5] We follow OECD index http://www.oecd.org/employment/protection

Table 16.1 Descriptive statistics by country

	Norway	Denmark	Finland	Sweden	Germany	UK	Sum
# observations	72	39	30	46	217	21	425
# studies	5	7	3	9	15	5	44
Programme effect:							
Negative	18	5	9	11	60	6	109
No effect	39	15	4	28	87	2	175
Positive	15	19	17	7	70	13	141
Programme:							
Training	23	1	12	6	46	3	91
Other programmes	0	0	0	0	26	0	26
Wage subsidies	21	0	8	8	12	6	55
Work practice	23	3	8	9	60	0	103
Employment	0	0	0	0	54	6	60
Intensified activation	5	35	2	23	19	6	90
Method:							
Diff-in-diffs	3	0	1	13	14	8	39
Timing of events	9	5	1	0	1	1	17
Matching	24	9	28	21	202	12	296
RCT	0	25	0	12	0		37
Other methods	36	0	0	0	0	0	36
Effect period:							
Short-term	24	30	16	23	99	13	205
Long-term	48	9	14	23	118	8	220
Outcome measure:							
Wage income	1	1	6	11	17	2	38
Employment	43	18	12	17	111	9	210
Unemployment	12	3	6	11	24	9	65
Welfare dependence	4	7	0	3	41	1	56
Education	12	10	6	4	24	0	56
By gender:							
All	30	20	30	43	56	13	192
Women	21	9	0	0	73	2	105
Men	21	10	0	3	88	6	128
By welfare dependence:							
All	72	30	30	37	84	0	253
With UB	0	2	0	6	6	21	35
With social security/ other benefits	0	7	0	3	127	0	137
By health conditions:							
All	69	30	30	38	207	21	395
Limited work capacity	3	9	0	8	10	0	30

(continued)

Table 16.1 (continued)

	Norway	Denmark	Finland	Sweden	Germany	UK	Sum
Period of evaluation							
Before 2004	72	5	28	41	82	21	177
From 2004 onwards		34	2	5	135	0	142
Youth unemployment	8.53	7.37	16.04	7.70	7.49	7.28	9.07
Rigidity of EPL*	2.30	2.13	2.37	2.67	2.70	1.16	2.22
Rigidity of TEC **	3.09	1.40	1.44	1.49	1.21	0.44	1.51
ALMP in percent of GDP	2.12	3.51	3.87	2.73	2.99	0.73	2.66

NB: macro variables are calculated by country over the period the programme was implemented. "Youth unemployment" and "public expenditure on ALMP in percent of GDP" are measured in percent. "Rigidity of the employment protection legislation" (EPL) and "rigidity of temporary employment contracts" (TEC) are indexed from 0 to 6 with 6 as most rigid (see http://www.oecd.org/employment/protection). The rest of the variables are measured in absolute values

effects, while Sweden has more observations with negative effects than with positive or non-significant effects. Germany has a uniform distribution of estimated programme effects. Norway has as many negative as positive effect estimates, but most effects are non-significant.

Germany is the country with most observations of all types of programmes, while Finland and the UK have a limited range of programme types. Norway and Germany have mostly evaluations of individual measures (training, wage subsidies, work practice, etc.), while Denmark and Sweden have most studies that analyse the effects of intensified activation. Furthermore, the Danish, Finnish and British studies focus mostly on short-term effects, while Norwegian and German studies have placed more emphasis on long-term effects of measures. Swedish studies have as many analyses of short as of long-term effects. Employment is the most frequently used outcome measure. While Norwegian, Danish and German studies have emphasized the study of effects by gender, the Finnish and Swedish studies have rarely been concerned about this. Most studies evaluate measures without conditioning on whether participants receive benefits or not. Germany is an exception to this pattern.

As regards the institutional and macroeconomic context, Table 16.1 shows that the Finnish youth initiatives have been implemented in a situation of substantially higher youth unemployment than in the other countries. The Nordic countries and Germany have similar institutional arrangements, with respect to employment protection legislation (EPL), the regulation of temporary contracts and the share of public expenditure on active labour market programmes (ALMP) as a percentage of gross domestic product (GDP). Britain has the least rigid regulations, both with regard to EPL and temporary contracts and less public expenditure on active labour market measures compared with the other countries included in the meta-analysis.

Descriptive statistics (not shown) relating to the sign of the impact indicate considerable variation: 41 percent of the observations indicate no impact, while 33 percent report a positive impact, and 26 percent report negative effects. The type of programme which has been evaluated the most is training, followed by work

experience, employment measures and intensified activation, in that order. Together, they account for more than 80 percent of all observations. Training, wage subsidies and intensified activation are the measures in our database where the majority of studies have found positive effects. Work practice has almost as many occurrences of negative as of positive effects. Employment programmes is the one with poorest results, mostly non-significant or negative effects.

Matching is the predominant method used and constitutes two-thirds of the number of observations. Matching provides relatively evenly distributed effects, while timing-of-event analysis produces the most positive effects. The relatively little-used methods in the residual category (OLS, structural models) give almost exclusively non-positive effects. Randomized trial, which is considered to be the best and most reliable method of evaluation, often produces non-significant effects. Most studies use job opportunity/employment probability as the outcome of interest. There are about as many evaluations of short-term as of long-term effects. Short-term effects provide more positive estimates than long-term effects, which are often non-significant.

16.3 Results

As mentioned above, we use a so-called ordered probit model to study how the estimated results vary with characteristics of the measure being evaluated, the conditions under which the programme was implemented and the estimation method applied to identify the impact. Our estimation strategy is to gradually add explanatory variables: we estimate model (1) including only dummy variables indicating programme types as regressors. Then in the model (2), we add variables related to programme characteristics. In model (3), we introduce dummy variables for countries and country-/time-specific youth unemployment rates. Finally, in model (4), we add institutional factors.

Tables 16.2 and 16.3 present estimates of the average marginal effect of the explanatory variables on the probability for negative and positive outcomes, respectively. That is, marginal effects are evaluated at the observed values of covariates and obtained from the corresponding ordered probit regressions.[6] The interpretation of the estimates is that, for example, in Table 16.2, model (1), when the effect of *work practice* is estimated to be 0.16, it means that relative to the benchmark programme, which is t*raining*, *work practice* is approximately 16% more likely to produce a negative effect. And since the number is statistically different from zero (as indicated by the stars), it can be interpreted as evidence that *work practice* has a significantly more negative impact than *training*. Similar interpretations can be obtained from estimates of positive marginal effects in Table 16.3. It is expected that

[6] Average marginal effect should not be confused with marginal effect at the mean. The first calculates the marginal effect for each case/observation in the data and calculate the means thereafter, rather than just the marginal effects at the mean value of each variable.

Table 16.2 Estimated marginal effects for obtaining a negative outcome

	Model (1)		Model (2)		Model (3)		Model (4)	
Programme (ref: training)								
Other programmes	0.2273	**	0.2573	**	0.2705	**	0.2648	**
Wage subsidies	−0.0219		−0.0642	*	−0.0797	**	−0.0829	**
Work practice	0.1632	**	0.1578	**	0.1438	**	0.1395	**
Employment programmes	0.2829	**	0.3637	**	0.3888	**	0.3796	**
Intensified activation	−0.0125		−0.0504		−0.0603		−0.0402	
Method (ref: OLS, other methods)								
Diff-in-diffs			−0.1768		0.0002		−0.0399	
Timing of events			−0.4111	**	−0.2817	**	−0.2937	**
Matching			−0.2569	**	−0.1437	**	−0.1450	
Randomized control experiments			0.0084		0.1752		0.2210	
Effect period (ref: short term)								
Long-term effect			−0.0286		−0.0290		−0.0291	
Outcome measure (ref: wage income)								
Employment			−0.0515		−0.0685		−0.0666	
Unemployment			0.0318		0.0333		0.0329	
Welfare dependency			0.0491		0.0283		0.0299	
Education			0.0229		−0.0118		−0.0105	
Separately by gender (ref: all)								
Women			−0.0669		−0.0810	*	−0.0830	
Men			−0.0812		−0.0997	*	−0.1024	
Separately by welfare subsidy (ref: all)								
Unemployment benefits			−0.1124	**	−0.1369	**	−0.1434	**
Social assistance, other subsidies			0.1628		0.1653		0.1413	
Separately for disabled			0.0678		0.0727		0.0690	
Period of evaluation (after 2003)			−0.2160		−0.3417	**	−0.3006	**
Country (ref: Norway)								
Denmark					0.0531		−0.1929	
Finland					−0.2587	**	−0.5090	*
Sweden					−0.0420		−0.3633	
Germany					0.0534		−0.3112	

(continued)

Table 16.2 (continued)

	Model (1)		Model (2)		Model (3)		Model (4)	
Great Britain					0.0089		0.2442	
Youth unemployment rate					0.0569	**	0.0365	
Rigidity of employment protection legislation							0.3681	
Rigidity of temporary employment regulation							−0.0738	
ALMP in percent of GDP							0.0442	
Pseudo R2	0.0468		0.1152		0.1276		0.1289	
No. of observations	425		425		425		425	

** indicates 5% significance level and * indicates 10% level. The pseudo R2 is obtained from ordered probit estimation

the estimates of marginal negative effects (Table 16.2) have the opposite sign to the estimates of marginal positive effects (Table 16.3), yet they are not a reflection of each other just with the opposite sign since no (significant) impact is also a possible outcome (trinomial dependent variable).

One first thing to notice in Tables 16.2 and 16.3 is that programme type is highly correlated with how successful programmes are in improving the labour market prospects of young people. First let us look at the marginal effects by programme type in model (1), in Tables 16.2 and 16.3. Model (1) in Table 16.2 shows that both work practice and employment measures are more likely to yield a higher probability of negative treatment effects relative to training programmes (reference category), as expressed by the significant p ositive p robabilities to produce negative marginal effects. The same picture can be seen in Table 16.3 In model (2) we add explanatory variables that control for programme-related characteristics. Interestingly, both the impact of wage subsidies and intensified activation programmes change considerably in size. However, it is only when we include country dummies and macro conditions that wage subsidies becomes the most successful of all programme types, with about 16% higher likelihood of producing a positive outcome compared to training programmes.

As regards the methods used in the estimations, we see that relative to the reference category (OLS and structural models), timing-of-event analysis and matching give positive effects (negative estimates in Table 16.2 and positive estimates in Table 16.3), while the use of quasi-experimental methods (mainly difference-in-differences methods) and experimental (randomized control trials) have no saying as to whether effects are positive or negative. This is robust to model specification. It is noteworthy that, when we include country dummies in model (3), the outcome varies dramatically with the method applied, reflecting that some countries have focused strongly in one particular method, like matching in Germany and experiments in Denmark.

Table 16.3 Estimated marginal effects for obtaining a positive outcome

	Model (1)		Model (2)		Model (3)		Model (4)	
Programme (ref: training)								
Other programmes	−0.2395	**	−0.2563	**	−0.2592	**	−0.2574	**
Wage subsidies	0.0367		0.1182		0.1495	**	0.1596	**
Work practice	−0.1898	**	−0.1826	**	−0.1666	**	−0.1640	**
Employment programmes	−0.2755	**	−0.3143	**	−0.3191	**	−0.3171	**
Intensified activation	0.0205		0.0892		0.1067		0.0682	
Method (ref: OLS, other methods)								
Diff-in-diffs			0.1362		−0.0002		0.0334	
Timing of events			0.5082	**	0.3750	**	0.3988	**
Matching			0.2255	**	0.1426	**	0.1425	*
Randomized control experiments			−0.0050		−0.1131		−0.1345	
Effect period (ref: short-term)								
Long-term effect			0.0316		0.0317		0.0317	
Outcome measure (ref: wage income)								
Employment			0.0604		0.0776		0.0754	
Unemployment			−0.0324		−0.0318		−0.0315	
Welfare dependency			−0.0487		−0.0273		−0.0288	
Education			−0.0236		0.0121		0.0108	
Separately by gender (ref: all)								
Women			0.0724		0.0858	*	0.0877	
Men			0.0901		0.1093	*	0.1121	
Separately by welfare subsidy (ref: all)								
Unemployment benefits			0.1848	**	0.2377	**	0.2461	**
Social assistance, other subsidies			−0.1633		−0.1634	*	−0.1412	
Separately for disabled			−0.0751		−0.0797		−0.0754	
Period of evaluation (after 2003)			0.2392	**	0.3747	**	0.3285	**
Country (ref: Norway)								
Denmark					−0.0549		0.1270	
Finland					0.5474	**	0.6047	**
Sweden					0.0506		0.3065	
Germany					−0.0553		0.2413	
Great Britain					−0.0099		−0.1031	

(continued)

Table 16.3 (continued)

	Model (1)		Model (2)		Model (3)		Model (4)	
Youth unemployment rate					−0.0624	**	−0.0398	
Rigidity of employment protection legislation							−0.4023	
Rigidity of temporary employment regulation							0.0807	
ALMP in percent of GDP							−0.0483	
Pseudo R2	0.0468		0.1152		0.1276		0.1289	
No. of observations	425		425		425		425	

** indicates 5% significance level and * indicates 10% level. The pseudo R2 is obtained from ordered probit estimation

The outcome studied, i.e. whether the dependent variable is income, employment or welfare dependence, does not seem to matter much for the finding reported. Nor does the time horizon of the effect of the programme, as shown by the non-significant differences in the probability of obtaining a more positive (or negative) outcome within the first years after completion of the programme relative to the impact more than a year later.

An interesting question is whether active labour market programmes are better for some groups than for others. Separate analyses by gender do not permit a clear interpretation; the estimand vary with the control variables included in the analysis. Programme effects also vary with whether the person receives unemployment benefits or not in the way that the effects seem to be more positive for unemployment benefit recipients, while there is an indication of the opposite when it comes to social assistance recipients. Needless to mention maybe, is that the limited number of studies by subgroup is likely to be causing low statistical power.

Another interesting result is that there seems to have been a learning process occurring over time. Results show that active labour market programmes seem to have a more positive impact in recent years than in the past, as expressed by a positive and significant estimate for programmes implemented during the last 10 years or so. This may indicate an improved efficiency in the way programmes are put together and implemented.

In model (3) we control for country-specific characteristics and indicators of economic/labour market conditions. We include dummy variables for each country and country-specific youth unemployment rates in the concomitant evaluation periods. Results indicate that, compared to Norway which is the reference category, Finland is the only country that has significantly more positive programme effects. However, this result needs to be interpreted with caution due to low number of studies from Finland. For the other Nordic countries, Germany and the UK there are no significant differences compared with Norway.

Results from model (3) also suggest that macroeconomic conditions matter. It turns out that when youth unemployment is high, programmes have less of a positive effect (more negative coefficients in Table 16.2 and less positive in Table 16.3). This is indicative that the effect of active labour market policies targeted at youth is procyclical: interventions work best during economic upturns, when the economy is recovering and there are available jobs for youth to take. This departs from findings in Card et al. (2015), which provides suggestive evidence that the effects of labour market programmes are better in recessionary markets. One potential reason is that our study focuses exclusively on countries where youth are differently affected by economic cycles than the population at large. Notably, when public expenditure on active labour market programmes is included, youth unemployment loses its explanatory power.

Finally, we include variables measuring institutional factors. We observe that there is no significant effect of institutional arrangements for how the active labour market programmes work. Such results are in line with findings in Kluve (2010). It should be noted, however, that there is little structural variation over time within countries such that the statistical uncertainty becomes large when all these variables are included simultaneously.

16.4 Summary and Discussion

We have conducted a comprehensive meta-analysis of youth labour market programmes based on a total of 44 studies, providing 425 observations, from six North European countries (Norway, Denmark, Sweden, Finland, Germany and the UK). An ordered probit method is used to estimate the importance of the factors that may affect the likelihood that youth improve their employability in the shorter or longer run. Our estimation results suggest that training programmes and wage subsidies generally give rise to more positive evaluation results than other measures. Work practice and employment in the public sector clearly underperform in quantitative evaluation studies.

The finding that evaluated effects of active labour market programmes do not depend on the outcome being evaluated is indicative that effects are robust in this respect. It is somewhat surprising that the timing-of-event analyses tend to produce more positive effects than other methods of analysis. One possible explanation is that timing-of-event analysis is a relatively new method that has been mostly used in recent years, mainly due to the increasing availability of comprehensive register data that permit the researcher to follow individuals over time in and outside the labour market. Because it is a relatively small sample of studies that use timing-of-event analysis, we believe that there may be some uncertainty associated with the interpretation of this result. On the other hand, traditional methods such as matching also show positive effects, while randomized control trials do not stand out from other methods. Since randomized control trials are regarded as the most reliable of all methods, and given that there are about as many studies from randomized trials

with significant positive and negative results, we conclude that programmes have actually both positive and negative effects. The fact that most findings are non-significant may to some extent reflect that the number of observations is often rather small in such controlled experiments.

Apart from Finland, where the active labour market policies seem to be more effective, there are not any significant differences across countries. We have not found any differences due to institutional factors either. These findings may be the result of small sample size and lack of variability between countries. It is noteworthy that Kluve (2010) found no impact of the institutional context either. Our results are in line with some of the findings in Card et al. (2010, 2015) and Kluve (2010) in other respects as well. Methodologically, they find that randomized trials do not give significantly different results than non-experimental methods. Our study reaches the same conclusion.

One weakness of meta-analyses often discussed is the so-called publication bias which is that studies that find significant results get more easily published. Since we do condition on the studies being published in scientific journals, the likelihood of publication bias is considerably reduced.

Ideally one would like to have more precise information than the sign of the effect. Impacts can be statistically significant but economically uninteresting if they are very small. Card et al. (2015) go to the great effort of calculating all impacts on employment in percent. If we were to pursue that line, the number of studies would be considerably reduced, since employment is only one of our measures. Nevertheless, it can be mentioned that they come to the conclusion that models based on the sign of the effect arrive at similar conclusions as those based on effect sizes.

It is challenging to perform a meta-analysis of this type. A caveat of our analysis is that there are relatively few evaluation studies of youth active labour market policies. This reduces the number of observations in our database. Thus, it becomes difficult to distinguish the importance of different factors from each other, such that the absence of statistical significance may be due to lack of data (degrees of freedom) rather than lack of "true" effect. For example, it is conceivable that the absence of significant differences in estimated effects related to institutional factors may be caused by too little variation in the institutional arrangements within each country. Thus, it becomes difficult to separate "country effects" from "institutional effects". Consequently, results must be interpreted with a degree of caution.

References

Bertola, G., Blau, F., & Kahn, L. (2007). Labor market institutions and demographic employment patterns. *Journal of Population Economics, 20*(4), 833–867.

Caliendo, M., & Schmidl, R. (2016). Youth unemployment and active labor market policies in Europe. *IZA Journal of Labor Policy, 5*, 1. https://doi.org/10.1186/s40173-016-0057-x.

Card, D., Kluve, J., & Weber, A. (2010). Active labour market evaluations: A meta-analysis. *The Economic Journal, 120*, F452–F477. https://doi.org/10.1111/j.1468-0297.2010.02387.x.

Card, D., Kluve, J., & Weber, A. (2015). *What works? A meta-analysis of recent active labor market programme evaluations*. NBER working paper no. 21431.
Greenberg, D. H., Michalopoulos, C., & Robins, P. K. (2003). A Meta-analysis of government-sponsored training programmes. *Industrial & Labor Relations Review, 57*(1), 31–53.
Jimeno, J. F., & Rodriguez-Palenzuela, D. (2003). *Youth unemployment in the OECD: Demographic shifts, labour market institutions, and macroeconomic shocks*. ECB working papers no. 155.
Kluve, J. (2010). The effectiveness of European active labor market programmes. *Labour Economics, 17*(6), 904–918.
Kluve, J., Puerto, S., Robalino, D., Romero, J. M., Rother, F., Stöterau, J., Weidenkaff, F., & Witte, M. (2016). *Do youth employment programs improve labor market outcomes? A systematic review.* IZA discussion paper no. 10263.
Puerto, O. S. (2007). *Labour market impact of youth: A meta-analysis of the Youth Employment Inventory*. Prepared for the World Bank http://www.youth-employment-inventory.org/downloads/7.pdf
Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. London: Routledge.

References of Studies Included in the Meta-analysis

Achatz, J., Fehr, S., Schels, B., & Wolff, J. (2012). *Ein-Euro-Jobs, betriebliche und schulische Trainingsmassnahmen: Wovon junge Arbeitslose im SGB II am meisten profitieren*. IAB-Kurzbericht 6, IAB Nuremberg.
Bell, B., Blundell, R., & Van Reenen, J. (1999). Getting the unemployed back to work: The role of targeted wage subsidies. *International Tax and Public Finance, 6*(3), 339–360.
Bennmarker, H., Grönqvist, E., & Öckert, B. (2013). Effects of outsourcing employment services: Evidence from a randomized experiment. *Journal of Public Economics, 98*, 68–84.
Bernhard, S., & Kruppe, T. (2012). Effectiveness of further vocational training in Germany. Empirical findings for persons receiving means-tested unemployment benefits. *Schmollers Jahrbuch: Zeitschrift für Wirtshafts- und Sozialwissenshaften, 132*(4), 501–526.
Bernhard, S., & Wolff, J. (2008). *Contracting out placement services in Germany. Is assignment to private providers effective for needy job-seekers?* Discussion paper 5, IAB Nuremberg.
Blundell, R., Costa Dias, M., Meghir, C., & van Reenen, J. (2004). Evaluating the employment impact of a mandatory job search programme. *Journal of the European Economic Association, 2*(4), 569–606.
Caliendo, M., & Künn, S. (2011). Start-up subsidies for the unemployed: Long-term evidence and effect heterogeneity. *Journal of Public Economics, 95*(3–4), 311–331.
Caliendo, M., Hujer, R., & Thomsen, S. (2008a). Identifying effect heterogeneity to improve the efficiency of job creation schemes in Germany. *Applied Economics, 40*(9), 1101–1122.
Caliendo, M., Hujer, R., & Thomsen, S. L. (2008b). The employment effects of job creation schemes in Germany: A micro econometric evaluation. In T. Fornby et al. (Eds.), *Modelling and evaluating treatment effects in econometrics*, *Advances in econometrics* (Vol. 21, pp. 381–428). Emerald Publishing Limited, registered in England with company number 03080506 and registered office Howard House, Wagon Lane, Bingley, West Yorkshire, BD16 1WA.
Caliendo, M., Künn, S., & Schmidl, R. (2011). Fighting youth unemployment: The effects of active labor market policies. IZA DP 6222.
Carling, K., & Larsson, L. (2005). Does early intervention help the unemployed youth? *Labour Economics, 12*(3), 301–319.
Costa Dias, M., Ichimura, H., & van den Berg, G. J. (2013). Treatment evaluation with selective participation and ineligibles. *Journal of the American Statistical Assocciation, 108*(502), 441–455.

Dahlberg, M., Johansson, K., & Mörk, E. (2009). *On mandatory activation of welfare recipients*. IZA DP No. 3947.

Dorsett, R. (2006). The new deal for young people: Effect on the labour market status of young men. *Labour Economics, 13*(3), 405–422.

Ehlert, C., Kluve, J., & Schaffner, S. (2012). Temporary work as an active labor market policy: Evaluating an innovative programme for disadvantaged youth. *Economics Bulletin, 32*, 1765–1773.

Engström, P., Hesselius, P., & Holmlund, B. (2012). Vacancy referrals, job search, and the duration of unemployment: A randomized experiment. *Labour, 26*(4), 419–435.

Forslund, A., & Nordström Skans, O. (2006). Swedish youth labour market policies revisited. *Quarterly Journal of Economic Research, 75*(3), 168–185.

Graversen, B. K., & Van Ours, J. C. (2008). How to help unemployed find jobs quickly: Experimental evidence from a mandatory activation programme. *Journal of Public Economics, 92*(10–11), 2020–2035.

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

Hall, C., & Liljeberg, L. (2011). *En jobbgaranti för ungdomar? Om Arbetsförmedlingens ungdomsinsatser*. IFAU- rapport 2011, 1.

Hämäläinen, K., & Ollikainen, V. (2004). *Differential effects of active labour market programmes in the early stages of young people's unemployment*. VATT research report 115, Helsinki.

Hämäläinen, K., & Tuomala, J.(2007). *Vocational labour market training in promoting youth employment.* VATT Discussion paper 432, Helsinki.

Hämäläinen, K., Hämäläinen, U., & Tuomala, J. (2014). *The labour market impacts of a youth guarantee: Lessons for Europe?* VATT WP 60, Helsinki.

Hardoy, I. (2003). *Effect evaluation: Methods and applications.* Accepted PhD thesis (dr.polit) at the University of Oslo, Norway.

Hardoy, I. (2005). Impact of multiple labour market programmes on multiple outcomes: The case of Norwegian youth programmes. *Labour, 19*(3), 425–467.

Hardoy, .I, Røed, K., Torp, H., & Zhang, T. (2006). Virker ungdomsgarantien? Søkelys på arbeidsmarkedet 1. Oslo.

Hohmeyer, K., & Wolff, J. (2012). A fistful of Euros: Is the German one-euro job workfare scheme effective for participants? *International Journal of Social Welfare, 21*, 174–185.

Huber, M., Lechner, M., & Wunsch, C. (2011). Do German welfare-to-work reduce welfare dependency and increase employment? *German Economic Review, 12*(2), 182–204.

Jensen, P., Rosholm, M., & Svarer, M. (2003). The response of youth unemployment to benefits, incentives and sanctions. *European Journal of Political Economy, 19*, 301–316.

Larsson, L. (2003). Evaluation of Swedish youth labor market programmes. *Journal of Human Resources, 38*(4), 891–927.

Maibom, J., Rosholm, M., & Svarer, M. (2012). Experimental evidence on the effects of early meetings and activation. IZA DP 6970, IZA Bonn.

Maibom, J., Rosholm, M., & Svarer, M.(2014). *Can active labour market policies combat youth unemployment?* IZA DP no. 7912, IZA Bonn.

Nivorozhkin, A., & Wolff, J. (2012). *Give them a break! Did activation of young welfare recipients overshoot in Germany? (A regression discontinuity analysis).* Conference paper, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2012: Neue Wege und Herausforderungen für den Arbeitsmarkt des 21. Jahrhunderts – Session: Labor Market Policy Evaluation. No. B08-V1.

Petrongolo, B. (2009). The long-term effect of job search requirements: Evidence from the UK JSA reform. *Journal of Public Economics, 93*(11–12), 1234–1253.

Rinne, U., Schneider, M., & Uhlendorff, U. (2011). Do the skilled and prime-aged unemployed benefit more from training? Effect heterogeneity of public training programmes in Germany. *Applied Economics, 43*, 3465–3494.

Rosholm, M., & Svarer, M. (2015). Kvantitativ evaluering af Brobygning til uddannelse. Rapport til Styrelsen for Arbejdsmarked og Rekruttering.

Svarer, M., Rosholm, M., Havn, L., & Høeberg, L. (2014). Evaluering af mentorindsats til unge uden uddannelse og job. Rapport til STAR, Rambøll.

van den Berg, G. J., Kjærsgaard, L., & Rosholm, M. (2012). *To meet or not to meet (your case worker) – That is the question.* IZA DP no. 6476, IZA Bonn.

van den Berg, G. J., Bozio, A., & Costa Dias, M. (2014a). *Policy discontinuity and duration outcomes.* IZA DP no. 8450.

van den Berg, G. J., Uhlendorff, A., & Wolff, J. (2014b). Sanctions for young welfare recipients. *Nordic Economic Policy Review, 1*, 177–208.

von Simson, K. (2012). Veier til jobb for ungdom som ikke har fullført videregående utdanning. *Søkelys på arbeidslivet, 29*(1–2), 76–96. Oslo.

Wolff, J., & Jozwiak, E. (2007). *Does short-term training activate means-tested unemployment benefit recipients in Germany?* Discussion paper 29, IAB Nuremberg.

Wolff, J., Popp, S., & Zabel, C. (2010). Ein-euro-jobs für hilfebedürftige Jugendliche. Hohe Verbreitung, geringe Integrationswirkung. *WSI-Mitteilungen, 63*(1), 11–18.

Zhang, T. (2016). Virker arbeidspraksis i ordinær virksomhet etter sitt formål? *Søkelys på arbeidslivet, 33*, 45–65.