

# What Works - Study into the Effects of Four Different Social Welfare Schemes: Pre-analysis Plan

Loek Groot, Stephanie Rosenkranz, Mark Sanders, Timo Verlaat  
Utrecht University School of Economics (U.S.E.)

December 21, 2018

## Abstract

Dissatisfaction with the current scheme of social welfare has led various Dutch cities to run field experiments testing different configurations of the scheme. Among them is the City of Utrecht, where the study goes by the name *Weten wat werkt* (English: What works) and includes more than 700 current social welfare recipients. For the duration of the trial job search requirements and welfare sanctions tied to benefit payments will be suspended to make room for more autonomy (treatment 1). Other aspects include a more tailor-made approach in supporting welfare recipients (treatment 2) and increasing financial incentives to take up paid work (treatment 3). The study is set up as a randomized controlled trial (RCT) with interventions running from June 2018 to September 2019. Outcomes of interest are differences in the uptake and quality of paid work, other forms of activation & participation, participants' health and well-being, participants' satisfaction with the new scheme, as well as their financial situation. In this document we set out our plan to analyse the experiment.

*JEL codes:* H530, I380, D910, C930

*Keywords:* welfare, labour, minimum income guarantee, income support, social assistance, policy evaluation

## 1 Introduction

In January 2015, a comprehensive social welfare reform introduced by the so-called Participation Act (*Participatiewet*) came into effect in the Netherlands. The reform also included a tightening of welfare regulations, increasing welfare conditionality. The new law caused some concern among municipalities, which are responsible for executing the scheme. Some local policymakers expected that further monitoring and control would prove to be bureaucratic and time-consuming. Others were calling into question the behavioural assumptions underlying a stricter scheme, namely: welfare recipients optimally trade-off cost and benefits of their actions and can therefore effectively be induced to comply by increasing the cost of non-compliance. Against this background, several municipalities<sup>1</sup>, among them the City of Utrecht, decided to experiment with different, more lenient configurations of the scheme.

In doing so, the Utrecht trial complements existing field research, which in the Netherlands has predominantly been focused on the effectiveness of active labour market programmes and stricter regimes (Bolhaar, Ketel, & van der Klaauw, 2018; de Koning, de Hek, Mallee, Rosing, & Groenewoud, 2014; van den Berg & van der Klaauw, 2006; van der Klaauw & van Ours, 2013), as well as interventions targeting psychological outcomes and (mental) health (Brenninkmeijer & Blonk, 2012; Schuring, Burdorf, Voorham, der Weduwe, & Mackenbach, 2009). The same holds true for field experiments in the welfare domain outside the Netherlands (Ashenfelter, Ashmore, & Deschênes, 2005; Behncke, Frölich, & Lechner, 2009; Graversen & van Ours, 2008). To the best of our knowledge, experimental evidence on reducing welfare conditionality is limited to trials in unemployment insurance (UI) in the U.S. in the 1990s, which showed that eliminating compliance activities increases UI receipt, but also improves reemployment quality (Johnson & Klepinger, 1994; Klepinger,

---

<sup>1</sup>Similar experiments take place in the Dutch cities of Amsterdam, Deventer, Epe, Nijmegen, Oss, Tilburg, and Wageningen. All experiments are supervised on the national level by the Dutch Ministry of Social Affairs and Employment and evaluated in close collaboration among the researchers involved.

Johnson, & Joesch, 2002).

With our experiment we aim to provide new insights into the working of welfare conditionality in social welfare. For the trial, we will vary central features of the Dutch status quo regime, such as (i) compliance requirements and sanctions, (ii) intensity of support, and (iii) retention of earnings, and measure impact on a broad range of outcome dimensions. That includes labour market outcomes, as well as documenting participants' job search behaviour, health, well-being, social participation, satisfaction with the scheme, and financial situation.

The remainder of this pre-analysis plan is structured as follows. Section 2 and 3 set out the policy context and the status quo social welfare regime, respectively. In Section 4 and 5 we introduce our research questions and interventions. Section 6 describes our experimental design. In section 7 and 8 we discuss threats to internal validity and our outcomes measures. We conclude by detailing our methods of analysis.

## 2 Policy context

It is instructive to briefly introduce the Dutch social welfare system. Social welfare (*bijstand*) in the Netherlands is a non-contributory transfer programme that provides monthly income support to poor households identified based on a means, and work test. The regime foresees a monthly transfer payment (also referred to as social welfare standard or *bijstandsnorm*) depending on the household composition.<sup>2</sup> In case income from work lies below the social welfare standard, partial payments can be collected to complement job earnings. On top of welfare benefits, claimants may be eligible for (means-tested) child, housing and healthcare allowances. The policy is designed as a temporary safety net aiming to deliver income support until recipients can provide for their own income again (mostly by finding employment). Whereas the legal framework (eligibility, benefit level, rules and

---

<sup>2</sup>As of January 1st, 2018, the social welfare standard is: maximum 992 EUR/month for a single-person household and 1,417 EUR/ month for a two-person household. Recipients receive extra budget for children living in the same household. Transfer payments are tied to the level of the statutory minimum wage: single-person households receive 70 percent of the net minimum wage, while two-person households receive 100 percent.

regulations) for social welfare is determined at the national level, local governments (municipalities and regional councils) are charged with the task of executing the scheme, that is, helping claimants to reintegrate into the labour market or participate in another way. Consequently, claimants apply for social welfare benefits in the municipality where they live and become clients of the local welfare agency, which often is a department at the municipality.

Our experiment takes place in Utrecht, which is the fourth largest city in the Netherlands and counts 347,574 inhabitants (1-1-2018). As of 31st December 2017, 10,124 households in Utrecht receive social welfare benefits, which makes up for 5.7 percent of all households. As a comparison, in the Netherlands as a whole, which counts around 17.1 million inhabitants (1-1-2018), 389,200 households receive social welfare benefits (6.0 percent of all households). In Utrecht, the department Work & Income (*Werk en Inkomen*) at the municipality is in charge of executing the social welfare scheme.

### 3 Status quo regime

In what follows we set out the main rules and obligations that exist for welfare recipients subject the status quo regime. We thereby distinguish between compliance activities and rules that concern the release of earnings on top of benefits. In general, claimants are obligated to provide the municipality with all the information that is necessary to assess their eligibility for social welfare benefits. That includes, amongst others, information on household composition, income and assets.

Furthermore, welfare recipients have to comply with job search and employment requirements. Those include:

1. Sending a certain amount of application letters each week.
2. Following activation and skill-development programmes.
3. Following education programmes (if required for a new job).
4. Signing up with a temporary employment agency.
5. Performing volunteer work.
6. Accepting all types of employment offered (including jobs that require a maximum

commuting time of three hours per day).

Recipients that fail to comply may be sanctioned by freezing or cutting their monthly payment. Benefit sanctions lie between 10 and 100 percent of monthly transfers. Exemptions apply to those who are incapacitated for work or face severe health problems. Caseworkers from the municipality oversee the re-integration process. The intensity of contact between the department and a claimant largely depends on the claimant's distance to the labour market (as assessed by the department). Claimants that are assumed to be close to finding paid work are in contact with the department more frequently.

If welfare recipients earn income on top of their benefits (e.g. through part-time work), most of that income is set off against the benefit payment. Under the status quo regulations recipients are allowed to keep 25 percent of their earnings up to a maximum amount of 202 EUR per month for a maximum period of in total six months. Recipients become ineligible for the benefit as soon as their earnings exceed the social welfare standard (see Section 2).

#### 4 Research questions

The aim of our experiment is to study the impact of different social welfare schemes on a broad range of outcome dimensions. That includes job search behaviour and labour market outcomes, as well as participants' health, well-being, social participation, satisfaction with the scheme, and financial situation. We are interested in the effect of three configurations of the scheme as opposed to the status quo regime. The first configuration targets compliance requirements and benefit sanctions and aims at reducing welfare conditionality. The theoretical foundations for reducing welfare conditionality are rooted in evidence from behavioural and social sciences. Insights from research into the scarce mind (Mani, Mullainathan, Shafir, & Zhao, 2013; Mullainathan & Shafir, 2013), reciprocal behaviour (Fehr & Schmidt, 2003) and motivation crowding (Frey & Jegen, 2001) lead us to believe that a non-coercive approach based on personal autonomy and trust might achieve better results. The second configuration targets the intensity of support and has been introduced by the Ministry of Social Affairs and Employment as an antipole to the first configuration. The

third configuration targets retention of earnings on top of benefits and aims at providing stronger financial incentives to take up paid work. Our four main research questions are:

1. **Status quo regime vs. exemption regime.** What is the effect of exempting welfare claimants from the usual compliance requirements to maintain their benefit and the benefit sanctions tied to those requirements?
2. **Status quo regime vs. intensive support regime.** What is the effect of stepping up the support that welfare claimants receive in order to reintegrate into the labour market?
3. **Exemption regime vs. intensive support regime.** What is the effect comparing a laissez-faire scheme to an intensive support scheme?
4. **Status quo regime vs. more generous earnings release.** What is the effect of decreasing the retention of income earned on top of benefits?

## 5 Interventions

This study knows three treatment conditions, one control condition and two reference groups. In the control condition, claimants are subject to the rules and regulation of the Dutch status quo social welfare regime as described in Section 3. In the three treatment conditions the rules and regulations are varied. The three treatments tested are:

**Treatment 1 (exemption).** Claimants are exempted from the usual obligations to maintain their benefit, such as applying for jobs, accepting employment or participating in activation programmes. Benefit sanctions for this group are eliminated. Participants in this group are thus given freedom of choice and full autonomy when it comes to finding paid work or participating in another way. The programmes and services the welfare department offers are still available to participants in that group, but they are offered on a demand-driven basis. At least once every six months the claimant is contacted by the department to inquire about the current status.

**Treatment 2 (intensive support).** All rules and regulations of the status quo regime

stay in place. However, claimants in this group receive extra support through tailor-made supervision and intensive mediation from the welfare department. On average, participants in this group have approximately twice as many contact moments with the department as control group participants.

**Treatment 3 (earnings release).** All rules and regulations of the status quo regime stay in place. However, claimants in this group are allowed to keep earnings on top of their benefits to a larger extent than under the current regulation. As mentioned before, under the status quo regime, claimants are allowed to keep 25 percent of additionally earned income up to a maximum of 202 EUR per month for a maximum period of in total six months. Participants in group 3 are subject to a more generous scheme. They are allowed to keep 50 percent of additionally earned income up to a maximum of 202 EUR per month for a maximum period of 16 months (the whole trial period).

Additionally, by making use of administrative data we follow two reference groups: (i) welfare claimants that were invited to take part in the study but did not enrol, and (ii) a randomly selected group of claimants that never received an invitation to participate in the trial.

## 6 Experimental design

### 6.1 Timing of the trial

The trial runs from March 2018, starting with recruitment and enrolment, until February 2020, when the final report is planned to be published. Our interventions are scheduled to start on June 1st, 2018 and run for 16 months in total, until September 30, 2019. Table 1 below sets out the time scheme of the trial.

### 6.2 Target group, sample and recruiting

The target population for the experiment consists of all welfare claimants in Utrecht, unless participating is a risk or not possible (see the Appendix for a complete list of exclusion

Table 1

*Time scheme of the experiment*

<b>Date</b>	<b>Experiment phase</b>	<b>Surveys</b>
March – May, 2018	Enrolment phase	1. March – May, 2018
June 1, 2018 – September 30, 2019	Intervention phase	2. January – February, 2019 3. August – September, 2019
Starting October 2019	Analysis, reporting	

criteria). Implementing the exclusion criteria around 8,100 claimants turn out to be eligible for the trial, which makes up for around 65 percent of the total population of welfare claimants in Utrecht. Participation in the experiment is voluntary, which is why recruitment is organised through a large-scale information and enrolment campaign. Participation has to be confirmed by signing an informed consent sheet. Dropping out of the experiment is possible at any stage.

### 6.3 Method of randomisation

We randomise at the individual level (claimants) applying a stratified randomisation design. Variables to form strata are subjects' distance to the labour market as assessed by the municipality (four categories also referred to as *arrangement* ranging from short to great distance) and the type of benefit subjects receive (two categories: single-person household benefit or two-person household benefit). Per stratum units will be assigned to each of the four groups (1 control and 3 treatment conditions) with equal probabilities. Misfits will be randomly allocated within each stratum. We will use the user-written RANDTREAT module in Stata to carry out the randomisation.

### 6.4 Method and instruments of data collection

Data is collected using administrative data as well as surveys. Administrative data sources are the municipality's register and Statistics Netherlands. The former will be used to collect monthly (pseudonymised) data on benefit receipt, outflow, contact moments,



re-integration programmes and instruments, benefit sanctions, freezes and fines, as well as additionally earned income. The municipality also possesses data on individual and household characteristics. From Statistics Netherlands we collect data on employment and income abstracted from social insurance records as well as data on the use of health services.

To complement information from administrative sources we collect survey data. The collection of survey data takes place at three points of time. We start with a pre-treatment measurement, which allows us to establish a baseline measure. Two follow-up measurements will take place half way and at the end of the trial. The different stages of measurement allow us to keep track of treatment effects over the course of the experiment.

We administer surveys using computer-assisted personal interviewing (CAPI) and computer-assisted web interviewing (CAWI). Only participants themselves are interviewed. In case of CAPI, trained interviewers visit participants at home. Interview schedules will be registered and interviewing locations can be verified using GPS tracking. Both CAPI and CAWI surveys are administered using Nfield survey software. Surveys are available in five different languages (Dutch, English, Turkish, Moroccan, Standard Modern Arabic). The field team consists of people speaking those languages fluently. Translations were provided by a professional translation bureau. We bench test and pilot our surveys prior to the beginning of the field period.

## 6.5 Power calculation

Our goal is to assign 200 individuals to each of the four groups. A power test shows that a sample of 400 (control + treatment condition) is sufficient to detect a medium sized effect (Cohen, 1988) of 0.3 SD (two-tailed) with an alpha of 5 percent and a power of 80 percent.

To determine power for binary outcome variables we perform a power test for comparing two independent proportions using the example of finding full-time employment. From the municipality's administrative records we know that under the status quo regime in the course of 1.5 years on average 6 percent of claimants leave the scheme because they find full-time employment. With 200 subjects per group we can detect a group difference of 8.5

percentage points.<sup>3</sup>

Central features of our experimental design are chosen with the intent to maximise statistical power (stratified randomisation design, measuring outcomes at baseline).

## 7 Threats to internal validity

### 7.1 Risk and treatment of noncompliance

Some subjects in our study that are assigned to a treatment condition might go untreated (one-sided noncompliance or failure-to-treat).<sup>4</sup> The main possible reasons for failure-to-treat in our setting are administrative failure and refusal to be treated. If participants refuse to be treated, they return to the status quo regime and are therefore go untreated. We keep track of such drop outs and will use that information to account for noncompliance in our analyses (see Section 9.1). Administrative failures are harder to keep track of, as caseworkers' deviation from the experimental protocol might go unnoticed. We use register data on contact moments, benefit sanctions and release of earnings to check for signs of administrative failure-to-treat and register any failures detected.

If necessary, we will adjust our analyses for noncompliance in a way that we can still obtain estimates for causal effects. A realistic goal is the estimation of the average intention-to-treat (ITT) effect and the complier average causal effect (CACE, sometimes also referred to as local average treatment effect). It is likely that some of our subjects will receive partial treatment, e.g. if they decide to refuse treatment while the experiment is running. We apply and report three different ways to account for partial treatment. First, we treat partially treated subjects as fully treated. Following that approach our estimated CACE moves towards zero and we obtain a conservative estimate. Second, we calculate upper and lower bounds for our CACE estimate. The lower bound is obtained by following the previous approach. The upper bound is obtained by treating partially treated subjects

---

<sup>3</sup>Note that power is lowest for this kind of operationalisation of this specific outcome variable.

<sup>4</sup>We assume that subjects are either treated according to their assigned treatment condition or go untreated, meaning that they are not treated according to another treatment condition. As we generally control access to treatment we do not expect two-sided noncompliance to pose a large threat.

as untreated. The two estimates bound the true CACE if the average treatment effect for partially treated subjects lies between zero and the CACE. Third, we account for partial treatment by assigning “partial credit”. A subject that is treated for  $x$  months out of  $y$  total months that the trial is running might be classified as  $d_i = x/y$ .

## 7.2 Risk and treatment of attrition

We can think of several reasons for missing data in our study. We might face item and survey nonresponse, in particular during follow-up waves, or incomplete administrative data. We can also think of missing outcomes following treatment noncompliance, in case subjects drop out of the study and refuse any further data collection. We follow a two-fold strategy to address missingness which includes ex-ante prevention measures and assessing ex-post whether missingness potentially confounds our results. To prevent missingness ex ante we implement a number of measures, of which the most important are: (i) following-up on subjects that left the scheme due to various reasons (e.g. relocation, finding work), (ii) collecting data from Statistics Netherlands, which is a reliable source for administrative data, (iii) incentivising survey participation among control group subjects with vouchers.

To diagnose the risk of attrition-related bias ex-post we follow a fourfold approach. First, we determine overall attrition rates. Second, we assess whether the magnitude of missingness is different for treatment and control conditions. Third, we check whether attrition subjects are different in terms of a broad range of baseline characteristics (e.g. gender, age, educational attainment, housing situation, distance to labour market). Lastly, we measure whether the baseline characteristics of attrition subjects in the treatment conditions are significantly different from those in the control condition. Following this approach, we are able to assess whether missingness threatens the symmetry between assigned experimental groups. If worrying levels of attrition are found, we will place bounds on our treatment effects, estimating the largest and smallest treatment effects that we would obtain if the missing information were replaced with the most extreme outcomes.

### 7.3 Risk and treatment of spillovers

We suppose that risk of spillover effects between households is low. With 800 participants, we aim at including a mere 6 percent of the total population of social welfare recipients in Utrecht. Furthermore, participants will be spread geographically across all ten city districts. To obtain more reliable information about potential spillovers between households we ask participants at midline if they know persons outside their own household participating in the trial.

In contrast, we have to acknowledge that there is a risk of spillover effects within households. This stems from the fact that randomisation takes place at the individual level, while some welfare recipients (28 percent of our target population) share a household and therefore receive benefits together. By design, participants sharing a household should be equally distributed across experimental groups.<sup>5</sup> The same holds true for participants with a partner in the same group, a different group or outside the experiment. As we observe household composition and group assignment we can easily confirm that assumption. If necessary we can use information on household composition and group assignment to condition on the group of the partner in our analyses.

### 7.4 Implementation checks

In case we are not able to find an effect we want to rule out implementation failure as a possible explanation. To prevent implementation failure we introduce a testing scheme, that helps us to assess early on if our independent variable has been effectively manipulated. For an overview see Table 2.

Firstly, we look at objective data, that is register data from the municipality. In case of successful treatment implementation we expect to see a difference in contact moments initiated by the municipality between the treatments. Treatment 2 (intensive support) should have the highest amount of such contact moments, whereas the lowest amount of such contact moments should be found in treatment 1 (exemption). We don't expect to see

---

<sup>5</sup>Keep in mind that the type of benefit (single-person household benefit or two-person household benefit) is one of our stratification variables.

a difference with regard to contact moments between treatment 3 (earnings release) and control. In case of successful implementation we should also observe no benefit sanctions for subjects in treatment 1 and a 75 percent tax rate on additionally earned income in all conditions except treatment 3.

Secondly, we look at self-reported survey data (participants' knowledge). We assess (i) whether participants know in which group they were placed and more specifically (ii) if they know which welfare rules apply to them. To monitor the process of treatment implementation we schedule biweekly meetings between the municipality and the research team. This allows us to detect and address problems with implementation early on.

## 8 Outcome measures

### 8.1 List of outcome variables

Our outcome measures concern (i) labour market outcomes, that is (part-time) employment, labour market earnings and labour contracts. Next to that, we look at (ii) job search, (iii) activation and participation, participants' (iv) health and (v) well-being, (vi) participants' satisfaction with the scheme, and (vii) participants' financial situation. Below we list all outcome variables and the respective data sources<sup>6</sup> per outcome group. The list of outcomes is followed by a list of main covariates, where we list relevant variables per covariate category.

1. Labour market outcomes
  - (a) Outflow from welfare\*
  - (b) Outflow from welfare to work\*
  - (c) Income from work\*
  - (d) Total income (work + benefits)\*
  - (e) Permanency of labour contract\*

---

<sup>6</sup>Outcome variables and covariates marked by an asterisk (\*) come from administrative data sources; all remaining outcome variables are included in our surveys.

Table 2

*Overview of implementation checks*

<b>Condition</b>	<b>Definition of condition</b>	<b>Implementation checks</b>
Treatment 1	Subject are exempted from reintegration obligations and the according sanctioning regime.	<ul style="list-style-type: none"> <li>• Check administrative data on contact moments.</li> <li>• Check administrative data on benefit sanctions.</li> <li>• Check participants' self-reported knowledge on exemption.</li> </ul>
Treatment 2	Subjects receive extra support through tailor-made supervision and intensive mediation from the welfare agency.	<ul style="list-style-type: none"> <li>• Check administrative data on contact moments (were subjects invited and did they show up to meetings?).</li> </ul>
Treatment 3	Subjects keep 50% of additionally earned income up to max. 202 EUR/month for max. 16 months.	<ul style="list-style-type: none"> <li>• Check administrative data on income.</li> <li>• Check participants' self-reported knowledge on the new rules to keep income.</li> </ul>
Control	Dutch status quo social welfare regime.	<ul style="list-style-type: none"> <li>• Check administrative data.</li> <li>• Check administrative data on benefit sanctions.</li> <li>• Check administrative data on income.</li> </ul>

2. Job search outcomes

*Job search index: weighted standardised average of variables 2a-b*

- (a) Job search intensity (past 4 weeks)
- (b) Applications sent (past 4 weeks)

3. Activation and participation outcomes

*Activation index: weighted standardised average of variables 3a-d*

- (a) Doing volunteer work
- (b) Providing informal care
- (c) Following training or education
- (d) Setting up an own business

4. Health outcomes

*Health index: weighted standardised average of variables 4a-d*

- (a) Self-reported general health (first item of SF-36)
- (b) Use of medicine targeting nervous system (ATC4-N)\*
- (c) Hospital visits\*
- (d) GP visits\*

5. Well-being outcomes

*Well-being index: weighted standardised average of variables 5a-g*

- (a) Psychological well-being scale (MHI 5)
- (b) Life satisfaction (WVS)
- (c) Eudemonic well-being (EU-SILC 2013 Meaning of Life)
- (d) Perceived stress (Elo, Leppänen, & Jahkola, 2003)
- (e) Social partaking (PASS social partaking)
- (f) Social contacts
- (g) Trust (WVS)

6. Client satisfaction outcomes

*Satisfaction index: weighted standardised average of variables 6a-c*

- (a) Satisfaction with welfare department
- (b) Satisfaction with welfare services scale
- (c) Trust in welfare department

7. Financial situation outcomes:

- (a) Share of people having money left at the end of the month

## 8.2 List of main covariates

1. Individual level

- (a) Gender\*
- (b) Age\*
- (c) Marital status\*
- (d) Educational level
- (e) Nationality\*
- (f) Years living in the Netherlands
- (g) Personality dimensions (Big Five)

2. Household level

- (a) Number of dependent children in household\*
- (b) Household composition\*
- (c) District\*

3. Welfare variables

- (a) Distance to labour market (*arrangement*)\*
- (b) Duration of benefit receipt\*

4. Methodological variables



- (a) Stratum
- (b) Interviewer ID and survey method
- (c) Survey language
- (d) Group assignment of partner
- (e) Baseline survey filled after group assignment

## 9 Methods of analysis

### 9.1 Basic specification

As we make use of an experimental design, identification of treatment effects is straight-forward. Our basic specification to capture the impact of a treatment is:

$$Y_{it} = \beta_{0t} + \beta_{1t}^x T_i^x + X_i \delta_t + \epsilon_{it} \quad (1)$$

where  $Y_{it}$  is the outcome of interest for individual  $i$ , measured at  $t$  time periods after the start of the intervention. In case of survey data we measure outcomes at  $t = M$  (midline) and  $t = E$  (endline), in case of administrative data we observe weekly and monthly outcomes depending on the outcome of interest.  $T_i^x$  is a treatment dummy and takes the value 0 for subjects assigned to control and the value 1 for subjects assigned to treatment  $x$  with  $x \in \{1, 2, 3\}$  indicating the different treatment groups. The omitted category are thus subjects assigned to control. When we compare treatment 1 to treatment 2,  $T_i^x$  takes the value 0 for subjects assigned to treatment 1 and the value 1 for subjects assigned to treatment 2. In that case subjects assigned to treatment 1 become the omitted category. The vector  $X_i$  contains individual and household background characteristics as well as methodological covariates (see Section 8.2 for a list of covariates).  $\epsilon_{it}$  is the error term. The parameter of interest is  $\beta_{1t}^x$ , which describes the effect of a new policy regime  $x$  at  $t$  time periods after the start of the intervention. We will also compare all experimental groups to our reference groups.

In case of noncompliance, Eq. (1) reveals the intent-to-treat effect (ITT). We can estimate the complier average causal effect (CACE) by either dividing the ITT by the

estimated proportion of compliers in the assigned treatment condition (see Section 7.1) or instrumenting actual treatment status by assigned treatment status in a two-stage least squares (2SLS) regression.

In case we observe outcomes at baseline, we perform ANCOVA and difference-in-differences analyses to improve statistical power. For the ANCOVA analysis we condition on the baseline level of the outcome of interest,  $Y_{i0}$ , and our specification takes the following form:

$$Y_{it} = \beta_{0t} + \beta_{1t}^x T_i^x + X_i \delta_t + \gamma_t Y_{i0} + \epsilon_{it} \quad (2)$$

Our difference-in-differences specification is:

$$Y_{it} = \beta_{0t} + \beta_{1t}^x T_i^x + \beta_{2t} A_t + \beta_{3t}^x T_i^x \times A_t + \epsilon_{i,t} \quad (3)$$

where  $A_t$  is a time dummy taking the value 1 for periods in which treatment occurs. In Eq. 3 the treatment effect of interest is described by  $\beta_{3t}^x$ .

## 9.2 Heterogeneous effects

We will test whether the effects we observe vary for subjects with different household and personal characteristics. All characteristics considered are observed at baseline. We test for heterogeneous effects along the following dimensions:

1. Gender
2. Age
3. Educational level
4. Household composition
5. Distance to labour market (*arrangement*)
6. Duration of benefit receipt
7. Capacity of work

Our basic specification to test for heterogeneous effects is:

$$Y_{it} = \beta_{0t} + \beta_{1t}^x T_i^x + \delta_t X_i + \beta_{3t}^x T_i^x \times X_i + \epsilon_{it} \quad (4)$$

where now  $X_i$  denotes the different household and personal characteristics for individual  $i$ .

### 9.3 Accounting for multiple inference

The aim of our study is to assess the working of different social welfare configurations on a broad scale, which is why we focus on multiple outcomes. Besides, some of our outcomes are measured using various survey questions that address different dimensions of a certain behaviour or concept. In our analysis we will account for the fact that we test multiple hypotheses in two ways. Firstly, to reduce the number of hypotheses tested, we will collapse the information from several outcome variables into summary indices (see Section 8.1). Our summary indices are a weighted average of the respective standardised outcomes and we preferably construct them following the procedure suggested by Anderson (2008). Secondly, we will adjust our  $p$ -values to account for multiple inference using the False Discovery Rate (FDR). In our analyses we will report both, naïve and adjusted  $p$ -values.

## 10 References

- Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), 1481–1495. doi: 10.1198/016214508000000841
- Ashenfelter, O., Ashmore, D., & Deschênes, O. (2005). Do unemployment insurance recipients actively seek work? Evidence from randomized trials in four U.S. States. *Journal of Econometrics*, 125, 53–75. doi: 10.1016/j.jeconom.2004.04.003
- Behncke, S., Frölich, M., & Lechner, M. (2009). Targeting Labour Market Programmes – Results from a Randomized Experiment. *Swiss Journal of Economics and Statistics*, 145(3), 221–268. doi: 10.1007/BF03399281
- Bolhaar, J., Ketel, N., & van der Klaauw, B. (2018). *Caseworker’s Discretion and the Effectiveness of Welfare-To-Work Programs* (IZA DP No. 11666).

- Brenninkmeijer, V., & Blonk, R. W. B. (2012). The effectiveness of the JOBS program among the long-term unemployed: a randomized experiment in the Netherlands. *Health Promotion International*, *27*(2), 220–229. doi: 10.1093/heapro/dar033
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Edition)*. London: L. Erlbaum Associates.
- de Koning, J., de Hek, P., Mallee, L., Rosing, F., & Groenewoud, M. (2014). *Uitkomsten en ervaringen experimenten netto-effectiviteit re-integratie*. Rotterdam/Amsterdam: SEOR, Epsilon Research, Regioplan.
- Elo, A.-L., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work, Environment & Health*, *29*(6), 444–451. doi: 10.5271/sjweh.752
- Fehr, E., & Schmidt, K. M. (2003). Theories of Fairness and Reciprocity: Evidence and Economic Applications. In M. Dewatripont, L. P. Hansen, & S. J. Turnovsky (Eds.), *Advances in Economics and Econometrics Theory and Applications, Eighth World Congress* (Vol. 1, pp. 208–257). Cambridge: Cambridge University Press.
- Frey, B. S., & Jegen, R. (2001). Motivation Crowding Theory. *Journal of Economic Surveys*, *15*(5), 589–611. doi: 10.1111/1467-6419.00150
- Graversen, B. K., & van Ours, J. C. (2008). Activating Unemployed Workers Works; Experimental Evidence from Denmark. *Economics Letters*, *100*(2), 308–310. doi: 10.1016/j.econlet.2008.02.016
- Johnson, T. R., & Klepinger, D. H. (1994). Experimental Evidence on Unemployment Insurance Work-Search Policies. *The Journal of Human Resources*, *29*(3), 695–717. doi: 10.2307/146249
- Klepinger, D. H., Johnson, T. R., & Joesch, J. M. (2002). Effects of Unemployment Insurance Work-Search Requirements: The Maryland Experiment. *Industrial and Labor Relations Review*, *56*(1), 3–22. doi: 10.1177/001979390205600101
- Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty Impedes Cognitive Function. *Science*, *341*(6149), 976–980. doi: 10.1126/science.1238041
- Mullainathan, S., & Shafir, E. (2013). *Scarcity – The True Cost of Not Having Enough*. London: Penguin Books.
- Schuring, M., Burdorf, A., Voorham, A. J., der Weduwe, K., & Mackenbach, J. P. (2009). Effectiveness of a health promotion programme for long-term unemployed subjects with health problems: a randomised controlled trial. *Journal of Epidemiology & Community Health*, *63*(11), 893–899. doi: 10.1136/jech.2008.080432

- van den Berg, G. J., & van der Klaauw, B. (2006). Counseling and Monitoring of Unemployed Workers: Theory and Evidence from a Controlled Social Experiment. *International Economic Review*, 47(3), 895–936. doi: 10.1111/j.1468-2354.2006.00399.x
- van der Klaauw, B., & van Ours, J. C. (2013). Carrot and Stick: How Re-Employment Bonuses and Benefit Sanctions Affect Exit Rates from Welfare. *Journal of Applied Econometrics*, 28(2), 275–296. doi: 10.1002/jae.1265

## Appendix

### List of exclusion criteria

Choosing the target population for the trial, we have to exclude claimants that are not subject to the regular scheme of rules and regulations given their specific circumstances. In what follows we list all exclusion criteria that are applied:

1. Claimants who are younger than 27 years old.
2. Claimants who receive social welfare benefits for less than 10 weeks at the start of intervention.
3. Claimants who reach retirement age during the experiment.
4. Claimants who receive unemployment insurance benefits from the Dutch Employee Insurance Agency (UWV).
5. Claimants who are subject to the Natural Persons Debt Rescheduling Act (*WSNP traject*).
6. Claimants who are part-time entrepreneurs and claimants who receive social welfare benefits for the self-employed (*Bbz* or *IOAZ-witkering*).
7. Claimants who are admitted to a healthcare institution.
8. Asylum status holders with integration obligations.