

Giving Welfare Recipients More Autonomy or More Support? Lessons From Six Social Experiments in the Netherlands

Timo Verlaat*, Alice Zulkarnain, Jonneke Bolhaar, Stephanie Rosenkranz, Peter Gramberg, Ruud Muffels, Mark Sanders, Loek Groot, Richard Rijnks, János Betkó, Egbert Jongen, Arjen Edzes, Henk-Wim de Boer, Viktor Venhorst, Niels Spierings

Draft version June 25, 2023; please do not circulate

Abstract

Welfare states commonly rely on activation policies to help the unemployed find paid work and provide a livelihood again. We study the employment, health, and financial effects of more lenient activation policies for welfare claimants in six randomized controlled trials in the Netherlands. Subjects (i) were given more autonomy, e.g., by eliminating job search requirements and obligatory caseworker meetings, (ii) received intensive counseling personalized to their needs and wishes, or (iii) could keep more of their earnings in addition to benefits. Results suggest that eliminating requirements neither improves nor harms the employment chances of claimants. Counseling and a larger earnings disregard show positive labor market effects at some sites. We find no evidence for health and financial effects.

JEL codes: C93, I38, J64

Keywords: Welfare reform, autonomy, unemployment, randomized controlled trial

A list of author affiliations is available in the Supporting Online Materials. *Corresponding author: Timo Verlaat. Email: T.L.L.Verlaat@cpb.nl. Remaining authors in random order.

The experiments discussed in this article were pre-registered at the AEA RCT Registry. A list of registry information is available in the Supporting Online Materials. We are grateful to all study participants, the Municipalities of Utrecht, Deventer, Wageningen, Groningen, Nijmegen, and Tilburg, and a scientific supervisory committee installed by ZonMW. The Utrecht experiment received financial support from the European Social Fund (2018EUSF2011696). The Nijmegen experiment was partly funded by an ESF-SITS grant (2017EUSF201587). Verlaat acknowledges support from the Dutch Research Council (406.16.538).

1 Introduction

Multiple welfare states worldwide have policies in place to help unemployed workers find paid work and achieve economic self-sufficiency while ensuring a subsistence income. Examples are unemployment insurance schemes, which provide a temporary income after job loss, and universal welfare schemes, which often offer a safety net of last resort. Nowadays, advanced economies spend a substantial share of their budget on unemployment and activation, totaling 1,7% of GDP on average in 2019 (OECD, 2023).

Standard activation policies impose requirements on benefit claimants, for example, job search requirements, participation in training programs, or meetings with caseworkers (Venn, 2012). These requirements are thought to counteract the negative work incentives provided by income support and increase the efficiency of job matches in the labor market. Typically, claimants' adherence to requirements is monitored, and benefits are temporarily cut in case of violations.

In recent years, unconditional forms of support, such as basic income or unconditional benefits, have gained prominence in policy debates and on research agendas. Examples reach from unconditional cash transfers in development aid (Baird et al., 2018; Banerjee et al., 2017, 2019) to experiments with a partial basic income in Finland (Hämäläinen et al., 2022) and a universal basic income in Germany (Bohmann et al., 2021). Also, concerns about the complexity of existing income support, which is often shown to be poorly understood (Chetty & Saez, 2013; Kleven, 2019), have fueled an interest in reducing conditionalities.

Still, little is known about the effects of policies favoring autonomy over control, particularly in the context of a developed country. We address this knowledge gap by analyzing six recent and comparable policy experiments in the Netherlands that tested such more lenient activation policies for welfare claimants in randomized controlled settings. The policies tested were: (i) an exemption from all requirements tied to benefit receipt, such as job search requirements and obligatory meetings with caseworkers; (ii) intensive counseling considering the needs and wishes of claimants; and (iii) a lower benefit reduction rate, allowing claimants to keep more of their earnings in addition to benefits.

All experiments targeted working-age people, who are (long-term) unemployed and, in principle, able to work but not eligible for unemployment insurance benefits. Like many other European countries, the Netherlands provides this target group with a subsistence income through welfare benefits. Payments equal 70-190% of the statutory monthly minimum wage,

depending on household composition.¹ Although welfare benefits do not expire, they are intended as temporary assistance until people can provide a livelihood again.

Consequently, we study the effects of the experiments on labor market outcomes. Our outcomes of interest are (i) the probability of entering *substantial employment* (working at least 12 hours a week) and (ii) the probability of becoming *economically self-sufficient*, i.e., independent from benefits due to paid work (working at least 27 hours a week). In addition, we assess the effects of the interventions on health and financial outcomes.

We compile data from six independent and pre-registered randomized controlled trials (RCTs) conducted between 2017 and 2019 with roughly 12,000 welfare claimants in total. The experiments were possible due to a government waiver that allowed temporary deviations from the current legislation for research purposes. The trials were planned and implemented by municipal welfare offices, which administer the national welfare scheme. We link experimental data from these trials to comprehensive register data from Statistics Netherlands and survey data collected before and during treatment. We analyze effects separately for each site instead of pooling data due to differences in experimental design and implementation across the six municipalities.

Our main finding is that relaxing requirements for welfare claimants does not seem to have a statistically significant effect on labor market outcomes. Effect estimates are mostly close to zero or moderately positive but none are statistically significant. At some sites, additional implementation data allows us to alleviate concerns that the absence of effects is related to a weak contrast between the treatment and the status quo regime. Second, we find evidence that both intensive counseling and stronger financial work incentives promote job finding when including small jobs. At one site, both treatments lead to a 60% higher probability of working at least 12 hours a week at endline (significant at the 5%-level). At another site, subjects assigned to counseling have a 70% higher chance of leaving welfare due to paid work (significant at the 10%-level), although this result is not robust to the use of alternative specifications and outcome measures. Third, for all three interventions, we find no evidence of effects on broader outcomes, such as health expenditures and problematic debts.

With this article, we expand the understanding of the type of policies that can increase the employment prospects of unemployed job seekers at the margins of the labor market. Our main contribution is describing the effects of policy interventions that have rarely been tested before.

¹ At the time of the experiments, benefit payments were to €992 (\$1,279 PPP) monthly for single-person households and €1,417 (\$1,827 PPP) monthly for two-person households.

Although the literature evaluating activation policies for the unemployed is extensive (Card et al., 2010; Kluve, 2010), only a few studies examined reduced monitoring or requirements (Bolhaar et al., 2020; Johnson & Klepinger, 1994; Klepinger et al., 2002; McVicar, 2008, 2010). In contrast to the treatment studied in this paper, none of those earlier interventions actively emphasized a switch from control to autonomy. Other related studies, like the 1970s negative income tax experiments in the U.S. and Canada, took place in institutional and family settings that are very different from today (Robins, 1985).

Likewise, the numerous (intensive) counseling programs subject to earlier work concern mandatory, one-size-fits-all schemes (Graversen & van Ours, 2008; Maibom et al., 2017; Markussen & Røed, 2016) rather than support personalized to the needs and wishes of claimants. Lastly, stronger financial work incentives in the form of earnings disregards have, to our knowledge, only been tested for single mothers on welfare (Knoef & van Ours, 2016) but not for a broader group of benefit claimants.

Our second contribution is methodological. Our research setting allows us to link experimental data from multiple sites and contexts to comprehensive and reliable administrative microdata. On the one hand, this feature allows us to study outcomes from different domains and to rule out attrition-related bias encountered in earlier studies relying solely on survey data (Ashenfelter & Plant, 1990). On the other hand, leveraging a randomized study design, we expand the experimental evidence base on labor market and welfare policies, where randomized designs are well-established but still represent a relatively small share of the literature.

The remainder of this article is organized as follows. Section 2 introduces the treatments implemented and Section 3 describes experimental methods. We present our results in Section 4, while Section 5 concludes.

2 The Treatments

The experiments tested three treatments. The first two treatments involved changes in supervision and requirements. The third treatment solely altered financial incentives to take up work. For each treatment, we now describe the basic design elements. Table 1 provides an detailed overview of treatment components, including differences between experimental sites.

The first treatment (*autonomy*) eliminated requirements tied to benefit receipt, such as obligations to actively look for work, accept suitable job offers, meet caseworkers, and participate in labor market activation programs. The treatment also suspended benefit sanctions, i.e., a temporary cut in benefits, and monitoring by the welfare office.

The second treatment (*counseling*) offered a more intensive supervision and coaching program, personalized to the needs and wishes of claimants. Core components were (i) more frequent meetings with a caseworker, (ii) a lower caseload among caseworkers delivering the treatment, and (iii) a demand-oriented provision of services.

The third treatment (*earnings release*) lowered the benefit reduction rate, i.e., the rate at which benefits would reduce in case of additional earnings. The treatment also extended the period during which the lower reduction rate applies. Under status quo legislation, claimants may keep 25% of earnings in addition to benefits up to a maximum of €202 (\$260 PPP) per month. This earnings disregard is applicable for a maximum of six months. The treatment increased the disregard to 50%, while the maximum of €202 per month remained in place. It also eliminated the time limit, such that claimants could keep a fraction of additional earnings for as long as the respective experiment lasted (15 to 27 months). Some experimental sites combined the first two treatments with the third treatment (Deventer, Nijmegen, and Tilburg).

At some sites, additional treatments were administered.² As these treatments are less comparable across sites, they lie outside the scope of this article.

² A description of these treatments can be found in the pre-registrations.

Table 1. Summary Treatments

	Utrecht	Deventer	Wageningen	Groningen	Nijmegen	Tilburg
Treatment groups						
Group 1	Autonomy	Autonomy + earnings release	Autonomy	Autonomy	Autonomy + earnings release	Autonomy + earnings release + work bonus
Group 2	Counseling	Counseling + earnings release	Counseling	Counseling	Counseling + earnings release	Counseling + earnings release + work bonus
Group 3	Earnings release		Earnings release	Earnings release		
Treatment components						
Autonomy treatment	No compliance requirements, contact with caseworker every six months, support if requested	No compliance requirements, contact with caseworker every six months, drafting plan of action, support if requested	No compliance requirements, contact with caseworker every six months, training self-reliance	No compliance requirements, contact with caseworker every six months, no support available	No compliance requirements, contact with caseworker every six months, support if requested	No compliance requirements, contact with caseworker every six months, training self-reliance
Counseling treatment	Permanent caseworker, more frequent contact, additional support instruments	Permanent caseworker, biweekly contact, drafting plan of action	Permanent caseworker, more frequent contact	Tailorized support by group of experienced experts	Tailorized support and intensive group coaching	Permanent caseworker, bimonthly contact, additional support instruments
Earnings release treatment	50% instead of 25% up to €202 per month	25% instead of 0% up to €202 per month	50% instead of 25% up to €202 per month	50% instead of 25% up to €202 per month	50% instead of 25% up to €202 per month	50% instead of 25% up to €202 per month
Work bonus treatment						Bonus of €2.400 for full-time job

3 Experimental Methods

3.1 Design

All experiments sampled subjects among claimants of welfare benefits at the respective sites. Roughly 60% of claimants were eligible for the experiments after excluding, for example, claimants in care facilities or claimants reaching the statutory pension age during the trial. We will refer to the eligible population as the *target population*. Eligibility criteria were largely harmonized across sites, with some exceptions described in Table 2 alongside other design elements.

Three experiments recruited a stock sample of claimants (Groningen, Utrecht, Nijmegen) and three experiments included both a stock sample and a sample of new inflow (Deventer, Tilburg, Wageningen). At some sites, recruitment took place in several rounds, creating cohorts of participants. The government waiver required participants to consent before being included in an experiment. Four experiments randomized a sample of consenting subjects, while two experiments randomized the entire target population prior to informed consent. The sample size varies between 752 (Utrecht), 780 (Tilburg), 366 (Nijmegen), and 410 (Wageningen) subjects for experiments using randomization after informed consent. The sample size at the other two sites (randomizing the whole target population) is 8,338 (Groningen) and 1,789 (Deventer) subjects.

All experiments used randomization at the individual level to assign subjects to treatment and control conditions. All sites used equal probabilities to assign subjects to experimental groups. One site (Utrecht) used a stratified randomization design with two strata. Experiments with a flow sample randomized new benefit claimants every two to four weeks for about one year. Randomization was executed by the researchers involved at all sites but Groningen and Nijmegen, where the implementation partner was responsible. Participants were informed about their group assignment two to three weeks before the start of treatment by the welfare office.

The experiments took place simultaneously, with start dates between October 2017 and June 2018 and an end date of December 31st, 2019. Depending on the site and on the cohort of participants, treatments lasted between 15 and 27 months.

3.2 Implementation

The treatments were implemented by the welfare offices of the respective cities. Table 2 presents details on the implementation. In the case of the *autonomy* and *counseling* treatment,

implementation included changes in supervision and monitoring. The implementation of the *earnings release* treatment merely required changes in administrative systems to modify the settlement of benefit payments with earnings in addition to benefits. Implementation protocols were in place to specify the delivery of the *autonomy* and *counseling* treatment. Regular meetings between researchers and welfare offices took place to monitor adherence to the protocol. Overall, no major deviations from protocol were detected.

At some sites, designated caseworkers were assigned to deliver a specific treatment. At other sites, caseworkers could deliver various treatments simultaneously, including the care as usual treatment. We acknowledge that both approaches come with advantages and disadvantages. In the first case, caseworkers are more likely to treat subjects according to protocol due to specializing in a treatment. On the downside, this approach often requires to select a group of dedicated caseworkers which introduces the risk of caseworker effects. In the second case, while caseworker effects are less likely, caseworkers may find it harder to maintain the contrast between different treatments.

Table 2. Summary Experimental Design

	Utrecht	Deventer	Wageningen	Groningen	Nijmegen	Tilburg
Eligibility requirements						
Duration spell	≥10 weeks	≥12 weeks for stock sample	None	≥26 weeks	None	None
Age	≥27 years	≥27 years	None	≥27 years	None	≥27 years
Randomization						
Sequencing	Invite, randomize	Randomize, invite	Invite, randomize	Randomize, invite	Invite, randomize	Invite, randomize
Method	Equal probabilities, stratified, by research team	Equal probabilities, by research team	Equal probabilities, by research team	Equal probabilities, by welfare office	Equal probabilities, by welfare office	Equal probabilities, by research team
Stratification variables	Household composition, distance to labor market	None	None	None	None	None
Number of subjects						
Invited	8,338	1,789	724	8,338	6,200	4,835
Randomized	752	1,789	410	8,338	366	780
Treatment 1	189	376	106	1,641	122	193
Treatment 2	188	384	98	1,620	122	194
Treatment 3	187	None	113	1,649	None	None
Default control group	188	397	93	1,635	122	202
Second control group	None	205	None	146	None	None
Other groups	None	427	None	1,647	None	191
Cohorts and timing						
Multiple starting dates	No	Yes	Yes	No	Yes	Yes
Cohort 1	Stock sample	Stock sample	Stock sample	Stock sample	Stock sample	Stock sample
Cohort 2	None	Stock sample	Biweekly inflow	None	Stock sample	Biweekly inflow
Cohort 3	None	New inflow	None	None	None	None
Start of treatment	June 2018	Oct. 2017, Feb. 2018, July 2018	Between Dec. 2017 and Sep. 2018	Nov. 2017	Dec. 2017, Apr. 2018	Between Nov. 2017 and Oct. 2018
End of treatment	Dec. 2019	Dec. 2019	Dec. 2019	Dec. 2019	Dec. 2019	Dec. 2019
Total treatment duration	19 months	17-27 months	16- 27 months	27 months	21-25 months	15-26 months

3.3 Data and Outcomes

Subjects at all experimental sites can be linked to comprehensive register data by Statistics Netherlands using unique personal identifiers. We measure labor market outcomes using social security records, which cover the entire population of employed workers in the Netherlands. The records detail earnings, hours worked, and the type of employment every month. We use two pre-agreed binary outcome variables based on hours worked, indicating *significant employment* (working more than 12 hours a week) and *economic self-sufficiency* (working more than 27 hours a week), respectively.³ The threshold for economic self-sufficiency is set at 27 hours, because working that amount of hours for the statutory minimum wage leads to ineligibility for welfare benefits.

To measure health outcomes, we collect information on healthcare expenditures, specifically for *visits to a general practitioner (GP)*. In the Netherlands, GPs are the first point of contact in getting healthcare when encountering physical or mental health problems. Expenditures for GP visits are covered by statutory health insurance. In contrast to labor market information, data on health expenditures are only available yearly.

To measure financial outcomes, we use data on defaulting on statutory health insurance premium payments, which serves as a proxy for *problematic debts*. Statistics Netherlands does not maintain a centralized debts registry. We choose defaulting on premium payments as a proxy, as it has been shown to correlate with having financial difficulties (Posthumus et al., 2019). Individuals enter the registry after being in arrears for six months. Data are available yearly.

We also use register data to collect information on individual and household background characteristics. This information includes age, gender, migration background (Dutch, western, non-western) the highest level of educational attainment (low, intermediate, high, unknown), and household composition (couple, couple with kids, single, single parent, unknown). To proxy welfare and labor market histories, we measure the duration of the current welfare spell and the cumulative labor market earnings in the nine to three months before the start of the experiments.

The longitudinal nature of the data allows us to observe and control for pre-treatment levels of the outcome variables. We harmonize our data such that treatments at all sites and for all

³ The outcome variables were determined before starting the data analysis in an agreement between the involved research institutes, the scientific advisory committee, the Dutch Ministry of Social Affairs and Employment and the Netherlands Bureau for Economic Policy Analysis. The outcome variables are not specified in detail in the pre-registration, as the registration took place before the agreement.

cohorts start in month zero and construct a panel dataset from 24 months before the start of treatment until the treatments stopped. For yearly data, we observe outcomes up to four years before the start of treatment.

For additional analyses presented in the Supporting Online Materials, we use survey data on, for example, self-reported health and well-being. All experiments conducted three waves of surveys (baseline, midline, endline), covering identical questions. Survey outcomes can be linked to administrative data using unique personal identifiers.

3.4 Integrity

Balance

We find that the treatment and control groups at all experimental sites are sufficiently balanced in terms of baseline individual and household characteristics. Table 5-Table 10 in the Supporting Online Materials present the results of balancing tests using a regression framework. We reject the equality of means of control and treatment groups at the 10%-level for some characteristics at some sites. However, we do so for less than 10% of the tested hypotheses, which suggests that detected differences occurred by chance.

Comparing cumulative earnings, we find significant differences in pre-treatment labor market attachment between the control and treatment groups at three sites, Deventer, Nijmegen, and Tilburg. In Deventer, the *autonomy* group has a 43% higher mean income than the control group; in Nijmegen, the average income in the *autonomy* group is 36% lower relative to the control group (both significant at the 10%-level). In Tilburg, average income prior to the experiment is 46% lower in the counseling group compared to the control group (significant at the 5%-level).

At two sites, Nijmegen and Tilburg, imbalances in ex-ante labor market attachment are also apparent in substantial differences regarding pre-treatment employment probabilities. In Nijmegen, the probability of significant employment is almost 60% lower in both treatment groups compared to the control group in several months pre-treatment (significant at the 5%-level). Similarly, the *counseling* group in Tilburg shows an almost 60% lower probability of significant employment than the control group before the start of treatment (significant at the 5%-level).

Due to these imbalances, we are cautious with causal statements when discussing effect estimates in Nijmegen and Tilburg for the outcome variables reported in this article. Although our main specification controls for pre-treatment labor market attachment and lagged outcome variables, differences between control and treatment groups may still be driven by

(unobserved) compositional effects to some extent. We are confident about causal claims at the other sites.

Compliance

Due to the voluntary nature of the experiments, not all subjects assigned to treatment were actually treated (one-sided noncompliance). Reasons for one-sided noncompliance are either non-response to the treatment invitation or active withdrawal before or during the experiment. Table 13 in the Supporting Online Materials shows compliance rates per experimental group and site.

We account for one-sided noncompliance by estimating and reporting both intention-to-treat effects (ITT) and local average treatment effects (LATE). We deem it unlikely that subjects assigned to control were treated because the welfare offices controlled access to treatment and implementation checks did not suggest otherwise. Likewise, there were no indications that treated subjects received a different treatment than the one assigned.

Attrition

As the register data of Statistics Netherlands covers the entire population of the Netherlands, we only face attrition in the rare cases of death or migration. Attrition rates for administrative data span 1-2% and are symmetric across experimental groups. For survey data, attrition is considerably higher due to nonresponse, reaching 20-40% at endline. We discuss survey attrition in more detail in Section 4 together with empirical results.

3.5 Empirical Strategy

We estimate treatment effects using a linear probability model, regressing an outcome of interest in month (year) t on treatment dummies and a set of control variables, including individual characteristics and pre-treatment outcome levels. Due to differences in experimental design and implementation, we estimate models for each site separately instead of pooling data.

Our main specification takes the following form:

$$y_i^t = \alpha^t + \sum_{m=1}^M \beta_m^t D_{im} + X_i' \gamma + \delta y_i^{t-x} + q + \varepsilon \quad (1)$$

where y_i^t denotes a respective outcome for subject i in month (year) t , and D_{im} denote treatment dummies depending on the treatments administered at the respective site. Our default reference group is the control group. As control variables, we include the background characteristics described in Section 3.3, captured by vector X_i' , and outcome levels in months

6, 18, and 24 before the start of treatment, denoted by y_i^{t-x} .⁴ We add covariates to increase the precision of our estimates and control for any imbalances observed at baseline (McKenzie, 2012). Lastly, q describes fixed effects for the calendar quarter in which treatments started and ε denotes the error term.

Our parameters of interest are β_m^t , which describe intention-to-treat effects (ITT), i.e., the average effect of assignment to treatment. ITT effects are informative in our setting as the treatments were designed as non-mandatory programs. In an additional analysis presented in the Supporting Online Materials, we estimate local average treatment effects (LATE) by instrumenting actual treatment with random treatment assignment.

For Groningen, we do not use the original control group as a reference group due to concerns about reactive behavior by this group. Employment chances in the original control group quickly improve after the start of treatment, a pattern that remains unobserved in all treatment groups and a second control group that received no invitation to participate in the experiment that was only followed administratively. Therefore, we use the second (uninformed) control group as a reference group and report effect estimates in relation to the original control group in the Supporting Online Materials.

To assess the sensitivity of our results, we use alternative specifications with different control variables or alternative operationalizations of outcome variables. Our alternative specifications either exclude lagged outcome variables or all control variables. To operationalize outcome variables differently, we define *significant employment* and *economic self-sufficiency* based on earnings instead of hours worked and employ *drug use* as a proxy for individuals' health. We use a slightly modified specification when estimating effects based on survey data, including the baseline value of the respective outcome variable and, where applicable, dummies for survey mode and language.

4 Results

4.1 Labor Market Outcomes

Table 3 presents employment effects for all six experiments three months before the end of the experiment to avoid capturing anticipation effects of the ending treatment. Panel A reports results for *economic self-sufficiency*, and Panel B for *significant employment*. Column (1) shows control group means, while Column (2)-(4) present effect estimates. In the lower section

⁴ For yearly data, we control for outcomes in 2014-2016. In Utrecht, we also include dummies for randomization strata.

of each panel, we report the results for the three sites, where causal claims come with greater uncertainty, either due to changing the default control group (Groningen) or baseline imbalances (Nijmegen and Tilburg).

We find no evidence that the *autonomy* treatment, which provided an exemption from requirements, affected employment probabilities. The point estimates in Utrecht, Deventer, and Wageningen are mainly close to zero and statistically insignificant. Moderately positive point estimates are also insignificant, with lower bounds of the 90% confidence interval reaching values of -1.3 (Utrecht, significant employment) and -3.1 (Wageningen, economic self-sufficiency). For Nijmegen, we find a statistically significant negative point estimate for significant employment. However, given that we observe differences in employment chances of similar magnitudes already pre-treatment, we believe that this estimate cannot be interpreted as the causal effect of treatment.

The absence of treatment effects may relate to a weak contrast between the autonomy treatment and the status quo regime, given that welfare claimants with long unemployment spells and low labor market prospects are commonly supervised less intensively. However, we can demonstrate a treatment contrast at two sites (Utrecht and Nijmegen) using data on contacts and employment services. We find that contacts with the welfare office roughly halved in the *autonomy* groups from roughly eight to four contacts a year on average. In addition, the data in Utrecht shows that hardly any subject in the *autonomy* group received employment services compared to one-fifth in the control group. Table 11 Table 12 in the Supporting Online Materials provides detailed results.

We find positive and statistically significant employment effects of *intensive counseling* at two sites. Subjects assigned to counseling have a 6.7 percentage point (60%) higher probability of significant employment in Utrecht ($p < .05$) and a 7.8 percentage point (70%) higher probability of economic self-sufficiency in Wageningen ($p < .10$) relative to the control group. For Nijmegen, we again find a sizeable and significant negative point estimate for significant employment, which we do not interpret as a causal effect for the reasons mentioned above.

Stronger financial work incentives in the form of an *earnings release* improved employment outcomes at one site, leading to a 6.6 percentage point (60%) higher probability of significant employment in Utrecht ($p < .05$) compared to the control group. Remember that the earnings release treatment in Utrecht was a single treatment and not combined with autonomy or counseling like in Deventer, Nijmegen, and Tilburg.

Our findings in Utrecht are largely robust to using alternative specifications, different cut-off values, and adjacent outcome variables. Point estimates slightly increase in size when

omitting lagged outcomes or individual covariates from the main specification (see Table 16 and Table 17 in the Supporting Online Materials) or when using a 9-hour threshold to determine significant employment (see Table 19). When using a 15-hour threshold, point estimates are smaller and only significant for *counseling*; when defining significant employment as earning more than 30% of the statutory monthly minimum wage, estimates are only significant for *earnings release* (see Table 18).

The result for Wageningen is sensitive to using alternative specifications and adjacent outcome variables. Point estimates decrease to 7.3 and 7.0 percentage points and are no longer statistically significant when estimating Eq.(1) without lagged outcomes or individual covariates, respectively. When defining economic self-sufficiency as earning more than 70% of the statutory monthly minimum wage, the point estimate reduces to 4.9 percentage points and is not statistically significant.

Table 3. Employment Effects at Endline

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Economic self-sufficiency (1/0)					
Utrecht	0.059 (0.236)	0.019 (0.026)	0.040 (0.027)	0.018 (0.025)	750
Deventer	0.087 (0.282)	0.009 (0.020)	0.007 (0.021)		1547
Wageningen	0.109 (0.313)	0.063 (0.048)	0.078* (0.047)	0.007 (0.043)	407
Groningen	0.082 (0.274)	-0.002 (0.020)	0.004 (0.020)	0.011 (0.020)	8191
Nijmegen	0.149 (0.357)	-0.033 (0.043)	-0.028 (0.043)		362
Tilburg	0.095 (0.294)	-0.005 (0.028)	-0.007 (0.027)		771
Panel B: Significant employment (1/0)					
Utrecht	0.118 (0.323)	0.040 (0.032)	0.067** (0.033)	0.066** (0.033)	750
Deventer	0.179 (0.383)	-0.009 (0.024)	-0.002 (0.025)		1547
Wageningen	0.293 (0.458)	-0.003 (0.063)	0.033 (0.063)	-0.065 (0.059)	407
Groningen	0.144 (0.351)	-0.032 (0.026)	-0.025 (0.026)	-0.012 (0.026)	8191
Nijmegen	0.405 (0.493)	-0.139** (0.055)	-0.137** (0.056)		362
Tilburg	0.220 (0.415)	-0.023 (0.037)	-0.034 (0.036)		771

Notes: Estimates of ITT effects three months before the end of treatment. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, cumulative earnings in nine to six months before treatment, and outcome values in months 8, 16, and 24 before treatment. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

4.2 Health and Financial Outcomes

Table 4 reports the effects on expenditures for a GP visit (Panel A) and problematic debts (Panel B) for all six experiments in 2019, the last year of treatment. The structure of the table follows the previous one, with control group means in Column (1), effect estimates in Column (2)-(4), and results for Groningen, Nijmegen, and Tilburg in each panel's lower section.

We find no evidence that any of the three treatments affected expenditures for a GP visit. The point estimates have positive and negative signs but are not statistically significantly different from zero. Even for larger estimates, like 15.7 euro in Wageningen, confidence intervals are too wide (90% CI: -15.7, 47.1) to confidently reject the null hypothesis. Our conclusions do not change when employing the use of prescription drugs as an alternative outcome variable instead of healthcare expenditures (see Table 20 in the Supporting Online Materials). The results also remain unchanged when omitting lagged outcomes and all individual covariates from our main specification (see Table 21 Table 22).

We find no evidence for effects on the probability of having problematic debts. In some cases, the point estimates are negative and sizeable relative to a low control group prevalence of 3-5% but not statistically significantly different from zero. In Utrecht, for example, the negative point estimate for counseling would correspond to a relative effect of -53%. Still, the 90% confidence interval entails positive values up to 0.5. Estimating alternative specifications does not change the results (see Table 21 Table 22).

Our conclusions about health effects do not change when considering outcomes reported in surveys. Table 15 in the Supporting Online Materials presents results for self-reported health and well-being, both scored on a scale from 0 to 10. We find no evidence for effects on self-reported health. Point estimates have positive and negative signs and reach magnitudes up to 0.2 SD (Wageningen, earnings release) but are not statistically significantly different from zero.

At one experimental site, Wageningen, results suggest adverse effects on self-reported well-being of 0.3-0.4 SD for *counseling* ($p < .10$) and *earnings release* ($p < .05$). However, we interpret these findings with caution due to potential attrition-based bias. Response rates at endline were low, reaching 55% in the control group and 46% and 56% in the *counseling* and *earnings release* group, respectively (see Table 14 in de Supporting Online Materials). In addition, effects for *counseling* are smaller and no longer statistically significant when omitting baseline self-reported well-being from the specification.

Table 4. Health and Financial Effects

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Expenditures GP visit (euro)					
Utrecht	227.491 (160.153)	-5.308 (14.071)	-10.292 (13.435)	-11.864 (13.441)	752
Deventer	223.842 (180.529)	-8.453 (10.627)	-1.732 (10.154)		1568
Wageningen	198.048 (153.855)	15.666 (19.061)	-11.059 (18.070)	-8.983 (18.187)	407
Groningen	189.665 (146.359)	-28.890 (18.118)	-21.530 (17.988)	-26.783 (17.971)	8236
Nijmegen	197.808 (147.519)	9.519 (16.674)	49.881 (31.947)		362
Tilburg	227.857 (153.560)	6.284 (14.934)	-3.915 (13.540)		775
Panel B: Problematic debts (1/0)					
Utrecht	0.038 (0.191)	-0.011 (0.016)	-0.020 (0.015)	-0.010 (0.017)	748
Deventer	0.056 (0.230)	-0.004 (0.015)	-0.018 (0.014)		1547
Wageningen	0.033 (0.179)	0.003 (0.025)	0.001 (0.025)	-0.016 (0.024)	407
Groningen	0.047 (0.211)	0.024 (0.015)	0.015 (0.015)	0.018 (0.015)	8162
Nijmegen	0.033 (0.180)	-0.009 (0.024)	0.021 (0.027)		361
Tilburg	0.030 (0.171)	0.033 (0.020)	0.004 (0.016)		769

Notes: Estimates of ITT effects in 2019. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, cumulative earnings in nine to six months before treatment, and outcome values in years 2014, 2015, and 2016. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

5 Conclusion

We analyzed the employment, health, and financial effects of three different activation policies for welfare claimants tested in six RCTs in the Netherlands. The treatments included (i) an exemption from all requirements tied to benefit receipt, e.g., job search requirements (*autonomy*), (ii) intensive counseling considering the needs and wishes of claimants (*counseling*), and (iii) a lower benefit reduction rate (*earnings release*). We used comprehensive register data from Statistics Netherlands to follow roughly 12,000 subjects during the 19-27 months of treatment. In addition, we used data available up to two years before the start of the experiments to assess the comparability of control and treatment groups.

Our results suggest that relaxing conditionalities neither improves nor harms the employment chances of welfare claimants. Effect estimates at endline are either close to zero or moderately positive but not statistically significant. Limited sample sizes prevent a more precise estimation of null effects. On a cautionary note, an absence of effects could also relate to limited contrast between the status quo regime and the *autonomy* treatment. However, we can confirm a treatment contrast at two sites (Utrecht and Nijmegen) using data on contacts and employment services. Given the substantial public expenditures devoted to monitoring welfare claimants, we consider the absence of significant negative effects an important finding.

Existing studies provide consistent evidence for the positive employment effects of intensive counseling. Our results confirm this finding for a counseling program personalized to the needs and wishes of claimants. We find positive and statistically significant effects at two of the six sites. At endline, subjects assigned to counseling had a 60% higher probability of working at least 12 hours a week in Utrecht. In Wageningen, we find a 70% higher probability of working at least 27 hours a week compared to the control group, although this result is not robust to using alternative specifications and adjacent outcome variables.

A more generous earnings release has previously been shown to stimulate employment among single mothers on welfare. Our results suggest that stronger financial incentives to take up work may also benefit a broader group of claimants. We find positive and statistically significant effects at one of the six sites. In Utrecht, subjects assigned to the earnings release were 60% more likely to work at least 12 hours a week than control subjects. Consistent with previous evidence, we find no evidence for effects on full-time employment and exits from welfare.

Lastly, we find no evidence for effects on health and financial outcomes, which could also be caused by data constraints, the treatment duration, or the size and type of the interventions.

Data on health and financial outcomes were only available on a yearly basis, which may have smoothed effects. Moreover, the duration of the interventions may have been too short or too narrow in scope to result in detectable health and financial effects.

The experiments reported in this study have limited sample sizes or low compliance rates, which prevents the detection of small treatment effects and the estimation of precise null effects. Future research will benefit greatly from larger samples to confirm the initial evidence provided by our study. Larger samples are also needed to study the heterogeneity of effects, e.g., for claimants with different labor market prospects. Lastly, further research is needed to uncover why an exemption from requirements does not harm employment chances, as hypothesized in classical economic theory.

References

- Angrist, J. D., & Imbens, G. W. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467–475. <https://doi.org/10.2307/2951620>
- Ashenfelter, O., & Plant, M. W. (1990). Nonparametric Estimates of the Labor-Supply Effects of Negative Income Tax Programs. *Journal of Labor Economics*, 8(1, pt. 2), S396–S415. <https://doi.org/10.1086/298255>
- Baird, S., McKenzie, D., & Özler, B. (2018). The effects of cash transfers on adult labor market outcomes. *IZA Journal of Development and Migration*, 8(22), 1–20. <https://doi.org/10.1186/s40176-018-0131-9>
- Banerjee, A., Hanna, R., Kreindler, G. E., & Olken, B. A. (2017). Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs. *The World Bank Research Observer*, 32(2), 155–184. <https://doi.org/10.1093/wbro/lkx002>
- Banerjee, A., Niehaus, P., & Suri, T. (2019). Universal Basic Income in the Developing World. *Annual Review of Economics*, 11, 959–983. <https://doi.org/10.1146/annurev-economics-080218-030229>
- Bohmann, S., Fiedler, S., Kasy, M., Schupp, J., & Schwerter, F. (2021). *Experimental evaluation of a Basic Income Pilot in Germany*. *AEA RCT Registry*, May 27. <https://doi.org/10.1257/rct.7734-1.0>
- Bolhaar, J., Ketel, N., & van der Klaauw, B. (2020). Caseworker’s discretion and the effectiveness of welfare-to-work programs. *Journal of Public Economics*, 183, 1–19. <https://doi.org/10.1016/j.jpubeco.2019.104080>
- Chetty, R., & Saez, E. (2013). Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients. *American Economic Journal: Applied Economics*, 5(1), 1–31. <https://doi.org/10.1257/app.5.1.1>
- Graversen, B. K., & van Ours, J. C. (2008). How to help unemployed find jobs quickly: Experimental evidence from a mandatory activation program. *Journal of Public Economics*, 92, 2020–2035. <https://doi.org/10.1016/j.jpubeco.2008.04.013>

- Hämäläinen, K., Verho, J., & Kanninen, O. (2022). Removing Welfare Traps: Employment Responses in the Finnish Basic Income Experiment. *American Economic Journal: Economic Policy*, 14(1), 501–522. <https://doi.org/10.1257/pol.20200143>
- Johnson, T. R., & Klepinger, D. H. (1994). Experimental Evidence on Unemployment Insurance Work-Search Policies. *Journal of Human Resources*, 29(3), 695–717. <https://doi.org/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. <https://doi.org/10.1177/00197939020560010>
- Kleven, H. (2019). *The EITC and the Extensive Margin: A Reappraisal*. mimeo. <https://doi.org/10.3386/w26405>
- Knoef, M., & van Ours, J. C. (2016). How to stimulate single mothers on welfare to find a job: Evidence from a policy experiment. *Journal of Population Economics*, 29(4), 1025–1061. <https://doi.org/10.1007/s00148-016-0593-0>
- Maibom, J., Rosholm, M., & Svarer, M. (2017). Experimental Evidence on the Effects of Early Meetings and Activation. *Scandinavian Journal of Economics*, 119(3), 541–570. <https://doi.org/10.1111/sjoe.12180>
- Markussen, S., & Røed, K. (2016). Leaving Poverty Behind? The Effects of Generous Income Support Paired with Activation. *American Economic Journal: Economic Policy*, 8(1), 180–211. <https://doi.org/10.1257/pol.20140334>
- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2), 210–221. <https://doi.org/10.1016/j.jdeveco.2012.01.002>
- McVicar, D. (2008). Job search monitoring intensity, unemployment exit and job entry: Quasi-experimental evidence from the UK. *Labour Economics*, 15(6), 1451–1468. <https://doi.org/10.1016/j.labeco.2008.02.002>
- McVicar, D. (2010). Does Job Search Monitoring Intensity Affect Unemployment? Evidence from Northern Ireland. *Economica*, 77(306), 296–313. <https://doi.org/10.1111/j.1468-0335.2008.00747.x>
- OECD. (2023). *Social Expenditure: Aggregated data (Edition 2020)*. *OECD Social and Welfare Statistics (database)*. <https://doi.org/10.1787/d3a45935-en>
- Posthumus, H. S., Doove, S., Strofer, W., & van Andel, W. (2019). *Verkenning geregistreerde problematische schulden*. Centraal Bureau voor de Statistiek. <https://www.cbs.nl/nl-nl/maatwerk/2019/12/verkenning-geregistreerde-problematische-schulden>
- Robins, P. K. (1985). A Comparison of the Labor Supply Findings from the Four Negative Income Tax Experiments. *Journal of Human Resources*, 20(4), 567–582.
- Venn, D. (2012). Eligibility Criteria for Unemployment Benefits: Quantitative Indicators for OECD and EU Countries. *OECD Social, Employment and Migration Working Paper No. 131*. <https://doi.org/10.1787/5k9h43kgkvr4-en>

Supporting Online Materials

Supporting Online Materials Text 1: List of Affiliations

Arjen Edzes	Hanze University of Applied Sciences & University of Groningen
Ruud Muffels	Tilburg University
Alice Zulkarnain	Netherlands Bureau for Economic Policy Analysis (CPB)
Egbert Jongen	Netherlands Bureau for Economic Policy Analysis (CPB) & Leiden University
Mark Sanders	Maastricht University
Timo Verlaat	Netherlands Bureau for Economic Policy Analysis (CPB) & Utrecht University School of Economics
János Betkó	Radboud University Nijmegen
Henk-Wim de Boer	Netherlands Bureau for Economic Policy Analysis (CPB)
Jonneke Bolhaar	Netherlands Bureau for Economic Policy Analysis (CPB)
Peter Gramberg	Saxion University of Applied Sciences
Loek Groot	Utrecht University School of Economics
Stephanie Rosenkranz	Utrecht University School of Economics
Richard Rijnks	University of Groningen
Niels Spierings	Radboud University Nijmegen
Viktor Venhorst	University of Groningen

Supporting Online Materials Text 2: Pre-registrations

Utrecht	AEARCTR-0003592
Deventer	AEARCTR-0004048
Wageningen	AEARCTR-0004170
Groningen	AEARCTR-0003618
Nijmegen	AEARCTR-0004250
Tilburg	AEARCTR-0004131

Tables S1: Balancing Checks

Table 5. Balancing Checks Utrecht

	Control mean (SD)	Autonomy	Counseling	Earnings release	N
	(1)	(2)	(3)	(4)	(5)
Age	46.771 (10.133)	0.126 (1.001)	-0.117 (1.033)	-0.607 (0.999)	752
Female (1/0)	0.489 (0.501)	-0.017 (0.051)	0.060 (0.051)	0.008 (0.052)	752
Lower education (1/0)	0.521 (0.501)	-0.090* (0.049)	-0.049 (0.048)	-0.097** (0.049)	752
Intermediate education (1/0)	0.229 (0.421)	0.029 (0.044)	0.038 (0.044)	0.033 (0.044)	752
Tertiary education (1/0)	0.191 (0.395)	0.072* (0.042)	0.027 (0.040)	0.065 (0.041)	752
Education unknown (1/0)	0.059 (0.235)	-0.011 (0.023)	-0.016 (0.022)	-0.001 (0.023)	752
No migration background (1/0)	0.351 (0.479)	0.048 (0.048)	0.024 (0.048)	0.025 (0.048)	752
Western migration background (1/0)	0.106 (0.309)	0.025 (0.033)	-0.021 (0.030)	0.022 (0.033)	752
Non-western migration background (1/0)	0.543 (0.500)	-0.074 (0.050)	-0.003 (0.050)	-0.047 (0.049)	752
Single hh (1/0)	0.511 (0.501)	0.004 (0.044)	-0.015 (0.045)	0.003 (0.044)	752
Single hh + kids (1/0)	0.261 (0.440)	-0.037 (0.043)	-0.002 (0.043)	-0.046 (0.042)	752
Cohabit hh (1/0)	0.053 (0.225)	0.024 (0.025)	0.003 (0.022)	-0.000 (0.023)	752
Cohabit hh + kids (1/0)	0.149 (0.357)	0.018 (0.023)	0.019 (0.025)	0.054** (0.025)	752
Other/unknown hh (1/0)	0.027 (0.161)	-0.011 (0.015)	-0.005 (0.016)	-0.010 (0.015)	752
Duration current spell (in months)	76.098 (74.273)	-6.660 (7.082)	-7.153 (6.773)	-1.529 (7.211)	752
Earnings before start (in euro)	370.697 (1159.632)	135.889 (125.562)	-11.094 (112.292)	85.145 (117.030)	752
Economic self-sufficiency t-8 (1/0)	0.000 (0.000)	0.021** (0.010)	0.021** (0.011)	0.005 (0.006)	752
Economic self-sufficiency t-16 (1/0)	0.000 (0.000)	0.021** (0.011)	0.022** (0.011)	0.000 (0.001)	749
Economic self-sufficiency t-24 (1/0)	0.022 (0.145)	-0.006 (0.014)	-0.000 (0.015)	-0.000 (0.015)	747
Significant employment t-8 (1/0)	0.048 (0.214)	0.014 (0.023)	0.015 (0.023)	0.037 (0.024)	752
Significant employment t-16 (1/0)	0.032 (0.177)	0.047** (0.023)	0.021 (0.021)	0.016 (0.020)	749
Significant employment t-24 (1/0)	0.043 (0.203)	0.009 (0.021)	0.021 (0.023)	0.026 (0.024)	747
Expenditures GP visit 2014 (euro)	167.304 (128.505)	-0.975 (13.087)	-11.416 (11.842)	0.889 (13.250)	729
Expenditures GP visit 2015 (euro)	163.892 (110.551)	-0.560 (12.201)	1.445 (11.106)	2.703 (11.859)	744
Expenditures GP visit 2016 (euro)	168.974 (105.555)	6.578 (11.126)	7.216 (11.917)	12.034 (11.972)	748
Problematic debts 2014 (1/0)	0.077 (0.268)	0.043 (0.031)	0.041 (0.031)	0.004 (0.029)	715
Problematic debts 2015 (1/0)	0.054 (0.227)	0.037 (0.027)	0.051* (0.028)	0.015 (0.025)	735
Problematic debts 2016 (1/0)	0.059 (0.227)	0.010 (0.027)	0.022 (0.028)	-0.005 (0.025)	746

Notes: Baseline differences in covariates and lagged outcomes between the control and treatment groups. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) shows coefficients on the treatment dummies when regressing the respective outcome on a set of treatment dummies and randomization strata fixed effects. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6. Balancing Checks Deventer

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Age	46.783 (10.177)	-0.842 (0.731)	-0.281 (0.736)		1584
Female (1/0)	0.554 (0.498)	-0.006 (0.036)	-0.033 (0.036)		1584
Lower education (1/0)	0.539 (0.499)	-0.004 (0.036)	-0.047 (0.036)		1584
Intermediate education (1/0)	0.295 (0.456)	-0.006 (0.033)	0.063* (0.034)		1584
Tertiary education (1/0)	0.098 (0.298)	0.027 (0.023)	-0.015 (0.021)		1584
Education unknown (1/0)	0.068 (0.252)	-0.016 (0.017)	-0.002 (0.018)		1584
No migration background (1/0)	0.524 (0.500)	0.002 (0.036)	-0.009 (0.036)		1584
Western migration background (1/0)	0.108 (0.311)	0.010 (0.023)	0.035 (0.024)		1584
Non-western migration background (1/0)	0.368 (0.483)	-0.013 (0.035)	-0.026 (0.034)		1584
Single hh (1/0)	0.418 (0.494)	0.021 (0.036)	0.083** (0.036)		1584
Single hh + kids (1/0)	0.257 (0.437)	-0.008 (0.031)	-0.029 (0.031)		1584
Cohabit hh (1/0)	0.106 (0.308)	0.011 (0.023)	0.003 (0.022)		1584
Cohabit hh + kids (1/0)	0.207 (0.405)	-0.025 (0.029)	-0.062** (0.027)		1584
Other/unknown hh (1/0)	0.013 (0.112)	0.001 (0.008)	0.005 (0.009)		1584
Duration current spell (in months)	63.999 (76.305)	-3.515 (5.218)	-5.041 (4.776)		1578
Earnings before start (in euro)	425.945 (1198.561)	192.174* (111.093)	67.380 (98.553)		1584
Economic self-sufficiency t-8 (1/0)	0.010 (0.100)	0.010 (0.009)	-0.011** (0.005)		1573
Economic self-sufficiency t-16 (1/0)	0.008 (0.088)	0.017* (0.009)	0.017* (0.009)		1546
Economic self-sufficiency t-24 (1/0)	0.011 (0.102)	0.018* (0.010)	0.004 (0.008)		1527
Significant employment t-8 (1/0)	0.051 (0.220)	0.015 (0.017)	0.004 (0.016)		1573
Significant employment t-16 (1/0)	0.049 (0.217)	0.027 (0.017)	0.027 (0.018)		1546
Significant employment t-24 (1/0)	0.063 (0.244)	0.025 (0.019)	0.016 (0.019)		1527
Expenditures GP visit 2014 (euro)	165.470 (109.463)	-3.608 (8.368)	-3.229 (8.518)		1493
Expenditures GP visit 2015 (euro)	165.499 (108.414)	0.977 (7.963)	1.857 (8.536)		1524
Expenditures GP visit 2016 (euro)	171.119 (121.107)	3.094 (8.363)	-6.535 (8.230)		1560
Problematic debts 2014 (1/0)	0.153 (0.361)	-0.011 (0.027)	-0.036 (0.025)		1481
Problematic debts 2015 (1/0)	0.151 (0.359)	-0.017 (0.026)	-0.031 (0.025)		1500
Problematic debts 2016 (1/0)	0.112	0.019	0.006		1531

Notes: Baseline differences in covariates and lagged outcomes between the control and treatment groups. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) shows coefficients on the treatment dummies when regressing the respective outcome on a set of treatment dummies and calendar quarter fixed effects. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Balancing Checks Wageningen

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Age	42.645 (12.722)	-4.916*** (1.797)	-1.494 (1.871)	-1.879 (1.712)	410
Female (1/0)	0.527 (0.502)	0.015 (0.071)	-0.038 (0.074)	-0.049 (0.071)	410
Lower education (1/0)	0.409 (0.494)	0.013 (0.070)	-0.057 (0.070)	-0.067 (0.068)	410
Intermediate education (1/0)	0.226 (0.420)	-0.037 (0.058)	0.027 (0.063)	0.019 (0.061)	410
Tertiary education (1/0)	0.226 (0.420)	0.027 (0.058)	-0.024 (0.058)	0.029 (0.059)	410
Education unknown (1/0)	0.140 (0.349)	-0.004 (0.050)	0.055 (0.054)	0.019 (0.051)	410
No migration background (1/0)	0.355 (0.481)	0.102 (0.069)	0.010 (0.070)	0.094 (0.069)	410
Western migration background (1/0)	0.108 (0.311)	-0.058 (0.039)	0.009 (0.046)	0.023 (0.045)	410
Non-western migration background (1/0)	0.538 (0.501)	-0.044 (0.071)	-0.019 (0.072)	-0.117* (0.070)	410
Single hh (1/0)	0.484 (0.502)	-0.015 (0.071)	-0.040 (0.071)	0.035 (0.069)	410
Single hh + kids (1/0)	0.204 (0.405)	0.012 (0.058)	-0.004 (0.059)	0.045 (0.059)	410
Cohabit hh (1/0)	0.043 (0.204)	0.037 (0.034)	0.035 (0.035)	-0.006 (0.029)	410
Cohabit hh + kids (1/0)	0.247 (0.434)	-0.022 (0.061)	0.010 (0.063)	-0.053 (0.058)	410
Other/unknown hh (1/0)	0.022 (0.146)	-0.012 (0.018)	-0.001 (0.021)	-0.021 (0.015)	410
Duration current spell (in months)	40.784 (50.772)	-1.285 (8.044)	5.518 (8.705)	-3.302 (7.120)	410
Earnings before start (in euro)	650.538 (1900.431)	-87.045 (219.173)	-141.859 (265.205)	-343.166 (229.084)	410
Economic self-sufficiency t-8 (1/0)	0.022 (0.147)	-0.003 (0.019)	-0.011 (0.022)	-0.005 (0.021)	400
Economic self-sufficiency t-16 (1/0)	0.024 (0.152)	0.040 (0.029)	-0.018 (0.020)	-0.015 (0.019)	379
Economic self-sufficiency t-24 (1/0)	0.013 (0.115)	0.033 (0.025)	-0.000 (0.019)	0.008 (0.018)	342
Significant employment t-8 (1/0)	0.077 (0.268)	-0.029 (0.034)	-0.033 (0.037)	-0.033 (0.036)	400
Significant employment t-16 (1/0)	0.106 (0.310)	0.020 (0.046)	-0.060 (0.040)	-0.065* (0.038)	379
Significant employment t-24 (1/0)	0.105 (0.309)	-0.013 (0.046)	-0.066 (0.042)	-0.024 (0.045)	342
Expenditures GP visit 2014 (euro)	163.856 (139.049)	-9.129 (20.694)	-2.884 (23.235)	-24.517 (19.651)	313
Expenditures GP visit 2015 (euro)	168.007 (194.038)	-4.312 (26.648)	0.187 (26.446)	-17.460 (24.754)	330
Expenditures GP visit 2016 (euro)	151.121 (187.971)	-5.033 (22.497)	-7.090 (22.816)	7.391 (23.186)	385
Problematic debts 2014 (1/0)	0.072 (0.261)	0.002 (0.043)	0.008 (0.046)	0.048 (0.048)	303
Problematic debts 2015 (1/0)	0.100 (0.302)	0.004 (0.049)	-0.065 (0.044)	-0.003 (0.050)	321
Problematic debts 2016 (1/0)	0.062	0.021	0.001	0.030	361

Notes: Baseline differences in covariates and lagged outcomes between the control and treatment groups. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) shows coefficients on the treatment dummies when regressing the respective outcome on a set of treatment dummies and calendar quarter fixed effects. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8. Balancing Checks Groningen

	Control mean (SD)	Autonomy	Counseling	Earnings release	N
	(1)	(2)	(3)	(4)	(5)
Age	42.645 (12.722)	-4.916*** (1.797)	-1.494 (1.871)	-1.879 (1.712)	410
Female (1/0)	0.527 (0.502)	0.015 (0.071)	-0.038 (0.074)	-0.049 (0.071)	410
Lower education (1/0)	0.409 (0.494)	0.013 (0.070)	-0.057 (0.070)	-0.067 (0.068)	410
Intermediate education (1/0)	0.226 (0.420)	-0.037 (0.058)	0.027 (0.063)	0.019 (0.061)	410
Tertiary education (1/0)	0.226 (0.420)	0.027 (0.058)	-0.024 (0.058)	0.029 (0.059)	410
Education unknown (1/0)	0.140 (0.349)	-0.004 (0.050)	0.055 (0.054)	0.019 (0.051)	410
No migration background (1/0)	0.355 (0.481)	0.102 (0.069)	0.010 (0.070)	0.094 (0.069)	410
Western migration background (1/0)	0.108 (0.311)	-0.058 (0.039)	0.009 (0.046)	0.023 (0.045)	410
Non-western migration background (1/0)	0.538 (0.501)	-0.044 (0.071)	-0.019 (0.072)	-0.117* (0.070)	410
Single hh (1/0)	0.484 (0.502)	-0.015 (0.071)	-0.040 (0.071)	0.035 (0.069)	410
Single hh + kids (1/0)	0.204 (0.405)	0.012 (0.058)	-0.004 (0.059)	0.045 (0.059)	410
Cohabit hh (1/0)	0.043 (0.204)	0.037 (0.034)	0.035 (0.035)	-0.006 (0.029)	410
Cohabit hh + kids (1/0)	0.247 (0.434)	-0.022 (0.061)	0.010 (0.063)	-0.053 (0.058)	410
Other/unknown hh (1/0)	0.022 (0.146)	-0.012 (0.018)	-0.001 (0.021)	-0.021 (0.015)	410
Duration current spell (in months)	40.784 (50.772)	-1.285 (8.044)	5.518 (8.705)	-3.302 (7.120)	410
Earnings before start (in euro)	650.538 (1900.431)	-87.045 (219.173)	-141.859 (265.205)	-343.166 (229.084)	410
Economic self-sufficiency t-8 (1/0)	0.022 (0.147)	-0.003 (0.019)	-0.011 (0.022)	-0.005 (0.021)	400
Economic self-sufficiency t-16 (1/0)	0.024 (0.152)	0.040 (0.029)	-0.018 (0.020)	-0.015 (0.019)	379
Economic self-sufficiency t-24 (1/0)	0.013 (0.115)	0.033 (0.025)	-0.000 (0.019)	0.008 (0.018)	342
Significant employment t-8 (1/0)	0.077 (0.268)	-0.029 (0.034)	-0.033 (0.037)	-0.033 (0.036)	400
Significant employment t-16 (1/0)	0.106 (0.310)	0.020 (0.046)	-0.060 (0.040)	-0.065* (0.038)	379
Significant employment t-24 (1/0)	0.105 (0.309)	-0.013 (0.046)	-0.066 (0.042)	-0.024 (0.045)	342
Expenditures GP visit 2014 (euro)	163.856 (139.049)	-9.129 (20.694)	-2.884 (23.235)	-24.517 (19.651)	313
Expenditures GP visit 2015 (euro)	168.007 (194.038)	-4.312 (26.648)	0.187 (26.446)	-17.460 (24.754)	330
Expenditures GP visit 2016 (euro)	151.121 (187.971)	-5.033 (22.497)	-7.090 (22.816)	7.391 (23.186)	385
Problematic debts 2014 (1/0)	0.072 (0.261)	0.002 (0.043)	0.008 (0.046)	0.048 (0.048)	303
Problematic debts 2015 (1/0)	0.100 (0.302)	0.004 (0.049)	-0.065 (0.044)	-0.003 (0.050)	321
Problematic debts 2016 (1/0)	0.062	0.021	0.001	0.030	361

Notes: Baseline differences in covariates and lagged outcomes between the control and treatment groups. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) shows coefficients on the treatment dummies when regressing the respective outcome on a set of treatment dummies. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9. Balancing Checks Nijmegen

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Age	45.525 (9.661)	-0.174 (1.242)	-2.546** (1.266)		366
Female (1/0)	0.533 (0.501)	-0.049 (0.064)	0.005 (0.064)		366
Lower education (1/0)	0.238 (0.427)	0.000 (0.055)	-0.031 (0.054)		366
Intermediate education (1/0)	0.369 (0.484)	-0.001 (0.062)	0.120* (0.063)		366
Tertiary education (1/0)	0.344 (0.477)	0.017 (0.061)	-0.080 (0.059)		366
Education unknown (1/0)	0.049 (0.217)	-0.016 (0.026)	-0.008 (0.027)		366
No migration background (1/0)	0.664 (0.474)	-0.016 (0.061)	-0.045 (0.062)		366
Western migration background (1/0)	0.090 (0.288)	0.024 (0.039)	0.009 (0.038)		366
Non-western migration background (1/0)	0.246 (0.432)	-0.008 (0.055)	0.035 (0.057)		366
Single hh (1/0)	0.582 (0.495)	-0.065 (0.064)	0.021 (0.064)		366
Single hh + kids (1/0)	0.287 (0.454)	0.049 (0.060)	-0.006 (0.058)		366
Cohabit hh (1/0)	0.025 (0.156)	0.016 (0.023)	-0.008 (0.018)		366
Cohabit hh + kids (1/0)	0.090 (0.288)	-0.017 (0.035)	-0.049 (0.032)		366
Other/unknown hh (1/0)	0.016 (0.128)	0.016 (0.020)	0.041* (0.024)		366
Duration current spell (in months)	54.263 (56.537)	2.430 (7.356)	-4.468 (7.060)		365
Earnings before start (in euro)	1127.492 (1938.575)	-413.846* (212.542)	-221.123 (228.992)		366
Economic self-sufficiency t-8 (1/0)	0.041 (0.200)	-0.033 (0.020)	-0.025 (0.022)		362
Economic self-sufficiency t-16 (1/0)	0.017 (0.128)	-0.016 (0.012)	-0.017 (0.012)		361
Economic self-sufficiency t-24 (1/0)	0.041 (0.200)	-0.025 (0.022)	-0.016 (0.023)		360
Significant employment t-8 (1/0)	0.157 (0.365)	-0.083** (0.041)	-0.090** (0.040)		362
Significant employment t-16 (1/0)	0.124 (0.331)	-0.058 (0.037)	-0.056 (0.038)		361
Significant employment t-24 (1/0)	0.157 (0.365)	-0.065 (0.042)	-0.090** (0.040)		360
Expenditures GP visit 2014 (euro)	183.346 (155.563)	-25.308 (17.013)	-15.523 (19.848)		356
Expenditures GP visit 2015 (euro)	161.331 (112.251)	0.524 (13.849)	1.289 (15.153)		362
Expenditures GP visit 2016 (euro)	184.374 (159.836)	-11.079 (17.814)	-4.450 (18.789)		362
Problematic debts 2014 (1/0)	0.077 (0.268)	0.084** (0.042)	0.068* (0.041)		354
Problematic debts 2015 (1/0)	0.084 (0.279)	0.078* (0.043)	0.042 (0.040)		356
Problematic debts 2016 (1/0)	0.066 (0.279)	0.035 (0.043)	0.027 (0.040)		359

Notes: Baseline differences in covariates and lagged outcomes between the control and treatment groups. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) shows coefficients on the treatment dummies when regressing the respective outcome on a set of treatment dummies and calendar quarter fixed effects. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 10. Balancing Checks Tilburg

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Age	44.515 (9.775)	1.436 (0.999)	-0.861 (1.015)		780
Female (1/0)	0.579 (0.495)	0.068 (0.049)	0.050 (0.049)		780
Lower education (1/0)	0.426 (0.496)	0.072 (0.050)	0.063 (0.050)		780
Intermediate education (1/0)	0.366 (0.483)	-0.076 (0.047)	-0.046 (0.048)		780
Tertiary education (1/0)	0.149 (0.356)	0.012 (0.036)	0.019 (0.036)		780
Education unknown (1/0)	0.059 (0.237)	-0.008 (0.023)	-0.036* (0.020)		780
No migration background (1/0)	0.644 (0.480)	-0.042 (0.049)	-0.056 (0.049)		780
Western migration background (1/0)	0.119 (0.324)	-0.015 (0.032)	-0.020 (0.031)		780
Non-western migration background (1/0)	0.238 (0.427)	0.057 (0.045)	0.076* (0.045)		780
Single hh (1/0)	0.545 (0.499)	-0.052 (0.050)	-0.082 (0.050)		780
Single hh + kids (1/0)	0.297 (0.458)	0.035 (0.047)	0.077 (0.047)		780
Cohabit hh (1/0)	0.069 (0.255)	-0.007 (0.025)	-0.008 (0.025)		780
Cohabit hh + kids (1/0)	0.069 (0.255)	0.024 (0.028)	0.028 (0.028)		780
Other/unknown hh (1/0)	0.020 (0.140)	0.001 (0.014)	-0.014 (0.011)		780
Duration current spell (in months)	56.846 (57.112)	6.777 (6.205)	1.963 (5.558)		780
Earnings before start (in euro)	573.896 (1486.992)	-130.833 (137.680)	-262.996** (124.295)		780
Economic self-sufficiency t-8 (1/0)	0.005 (0.070)	-0.005 (0.005)	-0.005 (0.005)		780
Economic self-sufficiency t-16 (1/0)	0.010 (0.100)	0.006 (0.011)	-0.005 (0.009)		776
Economic self-sufficiency t-24 (1/0)	0.015 (0.122)	-0.005 (0.011)	0.011 (0.014)		770
Significant employment t-8 (1/0)	0.094 (0.293)	-0.027 (0.027)	-0.053** (0.025)		780
Significant employment t-16 (1/0)	0.065 (0.247)	0.013 (0.026)	0.003 (0.025)		776
Significant employment t-24 (1/0)	0.066 (0.248)	0.002 (0.025)	0.023 (0.027)		770
Expenditures GP visit 2014 (euro)	164.944 (122.985)	-8.595 (11.485)	-8.013 (12.586)		756
Expenditures GP visit 2015 (euro)	171.044 (135.217)	-13.811 (12.795)	-9.527 (12.375)		770
Expenditures GP visit 2016 (euro)	167.786 (115.994)	3.230 (12.559)	6.648 (12.409)		777
Problematic debts 2014 (1/0)	0.191 (0.394)	0.000 (0.040)	0.044 (0.042)		747
Problematic debts 2015 (1/0)	0.185 (0.389)	-0.019 (0.039)	0.029 (0.041)		758
Problematic debts 2016 (1/0)	0.136 (0.389)	0.032 (0.039)	0.043 (0.041)		770

Notes: Baseline differences in covariates and lagged outcomes between the control and treatment groups. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) shows coefficients on the treatment dummies when regressing the respective outcome on a set of treatment dummies and calendar quarter fixed effects. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Tables S2: Implementation Checks and Compliance Rates

Table 11. Monthly Contacts and Employment Services Provided in Utrecht

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Monthly contacts with the welfare office					
Utrecht	0.69 (0.89)	-0.33*** (0.07)	0.62*** (0.09)	0.11 (0.09)	708
Panel B: Receiving any employment service (1/0)					
Utrecht	0.20 (0.40)	-0.12*** (0.04)	0.13*** (0.05)	0.06 (0.04)	673
Panel C: Receiving any job search assistance (1/0)					
Utrecht	0.14 (0.34)	-0.11*** (0.03)	-0.01 (0.04)	0.04 (0.04)	673

Notes: Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) shows coefficients on the treatment dummies when regressing the respective dependent variable on a set of treatment dummies and randomization strata fixed effects. Contacts are average monthly contacts per person initiated by the welfare office (in-person, digital, and by phone) during the 19 months of treatment. The dummy variables indicating employment services take the value 1 if the subject received services at least once in the first 16 months of treatment. Data on services is confined to subjects that had not withdrawn from the experiment by month 16. *** p<0.01, ** p<0.05, * p<0.1.

Table 12. Monthly Contacts in Nijmegen

	Control mean (1)	Autonomy mean (2)	Counseling mean (3)	N (5)
Panel A: Monthly contacts with the welfare office				
Nijmegen	0.63	0.27	1.16	

Notes: Column (1)-(3) report group means. Contacts are average monthly contacts per person in 2018 with the employment services of the welfare office (in-person, digital, and by phone). The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Nijmegen.

Table 13. Compliance Rates Across Experimental Groups

	Autonomy (1)	Counseling (2)	Earnings release (3)
Utrecht	89.4%	85.6%	85.6%
Deventer	26.1%	18.2%	
Wageningen	81.1%	84.7%	85.8%
Groningen	10.0%	6.4%	7.8%
Nijmegen	86.1%	80.3%	
Tilburg	76.2%	72.7%	

Note: Share of randomized subjects that actually received the assigned treatment per treatment group. Partially treated subjects counted as non-treated. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg.

Tables S3: Results Survey Data

Table 14. Survey Response Rates Across Experimental Groups

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Response Baseline Survey (0/1)					
Utrecht	0.920 (0.272)	0.001 (0.027)	-0.021 (0.030)	-0.021 (0.030)	752
Deventer	0.234 (0.424)	0.042 (0.031)	-0.009 (0.030)		1584
Wageningen	0.624 (0.487)	0.172*** (0.063)	0.065 (0.069)	0.160** (0.064)	412
Nijmegen	0.795 (0.405)	0.140*** (0.043)	0.097** (0.047)		366
Tilburg	0.673 (0.470)	0.005 (0.047)	-0.013 (0.047)		780
Panel B: Response Endline Survey (0/1)					
Utrecht	0.782 (0.414)	0.000 (0.042)	-0.074 (0.045)	-0.024 (0.043)	752
Deventer	0.179 (0.384)	0.027 (0.029)	-0.074*** (0.025)		1584
Wageningen	0.548 (0.500)	-0.009 (0.071)	-0.089 (0.073)	0.012 (0.070)	412
Nijmegen	0.680 (0.468)	0.058 (0.058)	-0.143** (0.062)		366
Tilburg	0.366 (0.483)	0.265*** (0.049)	0.244*** (0.049)		780

Notes: Differences in baseline and endline survey response rates between the control and treatment groups. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) shows coefficients on the treatment dummies when regressing a dummy for survey response on a set of treatment dummies and calendar quarter fixed effects. Robust standard errors are in parentheses. The model in Utrecht includes randomization strata fixed effects. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. We do not report results for Groningen, as there is no survey data available for the chosen reference group. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 15. Health and Well-being Effects at Endline

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Self-reported general health (0-10)					
Utrecht	3.134 (2.556)	0.072 (0.225)	0.264 (0.233)	-0.229 (0.230)	566
Deventer	3.786 (2.472)	-0.246 (0.328)	-0.378 (0.341)		232
Wageningen	4.216 (2.263)	-0.356 (0.365)	-0.060 (0.427)	-0.411 (0.383)	219
Nijmegen	4.217 (2.340)	-0.077 (0.283)	-0.119 (0.325)		237
Tilburg	3.699 (2.433)	0.021 (0.264)	-0.116 (0.258)		399
Panel B: Self-reported well-being (0-10)					
Utrecht	6.120 (2.083)	0.111 (0.207)	0.154 (0.215)	0.035 (0.211)	566
Deventer	6.293 (1.815)	0.029 (0.287)	-0.043 (0.291)		232
Wageningen	7.314 (1.609)	-0.453 (0.297)	-0.549* (0.322)	-0.728** (0.302)	219
Nijmegen	6.723 (1.447)	0.050 (0.198)	0.214 (0.183)		237
Tilburg	6.331 (1.849)	0.096 (0.222)	0.236 (0.213)		400

Notes: Estimates of ITT effects for the endline survey. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating a modified version of Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites. We do not report results for Groningen, as there is no survey data available for the chosen reference group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, cumulative earnings in nine to six months before treatment, the baseline value of the outcome variable, and dummies for survey mode and language, where applicable. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Tables S4: Results Sensitivity Analyses

Table 16. Employment Effects at Endline: No Lagged Controls

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Economic self-sufficiency (1/0)					
Utrecht	0.059 (0.236)	0.020 (0.026)	0.043 (0.027)	0.019 (0.025)	750
Deventer	0.087 (0.282)	0.013 (0.020)	0.010 (0.021)		1547
Wageningen	0.109 (0.313)	0.068 (0.049)	0.073 (0.048)	0.006 (0.045)	407
Groningen	0.055 (0.229)	-0.002 (0.020)	0.003 (0.020)	0.012 (0.020)	8191
Nijmegen	0.149 (0.357)	-0.043 (0.042)	-0.036 (0.043)		362
Tilburg	0.095 (0.294)	-0.006 (0.028)	-0.011 (0.027)		771
Panel B: Significant employment (1/0)					
Utrecht	0.118 (0.323)	0.052 (0.033)	0.077** (0.034)	0.076** (0.033)	750
Deventer	0.179 (0.383)	-0.007 (0.024)	0.004 (0.025)		1547
Wageningen	0.293 (0.458)	-0.008 (0.064)	0.017 (0.063)	-0.070 (0.060)	407
Groningen	0.138 (0.346)	-0.030 (0.026)	-0.025 (0.026)	-0.010 (0.026)	8191
Nijmegen	0.405 (0.493)	-0.141*** (0.054)	-0.149*** (0.056)		362
Tilburg	0.220 (0.415)	-0.023 (0.037)	-0.031 (0.036)		771

Notes: Estimates of ITT effects three months before the end of treatment. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating a modified version of Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, and cumulative earnings in nine to six months before treatment. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 17. Employment Effects at Endline: No Controls

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Economic self-sufficiency (1/0)					
Utrecht	0.059 (0.236)	0.024 (0.026)	0.041 (0.027)	0.026 (0.026)	750
Deventer	0.087 (0.282)	0.030 (0.022)	0.021 (0.022)		1553
Wageningen	0.109 (0.313)	0.076 (0.050)	0.070 (0.051)	0.015 (0.046)	407
Groningen	0.055 (0.229)	-0.004 (0.020)	0.005 (0.020)	0.013 (0.020)	8193
Nijmegen	0.149 (0.357)	-0.048 (0.043)	-0.041 (0.043)		363
Tilburg	0.095 (0.294)	-0.015 (0.029)	-0.022 (0.028)		771
Panel B: Significant employment (1/0)					
Utrecht	0.118 (0.323)	0.062* (0.034)	0.078** (0.036)	0.086** (0.035)	750
Deventer	0.179 (0.383)	0.022 (0.028)	0.022 (0.029)		1553
Wageningen	0.293 (0.458)	-0.001 (0.066)	0.016 (0.067)	-0.077 (0.062)	407
Groningen	0.138 (0.346)	-0.037 (0.030)	-0.029 (0.030)	-0.014 (0.030)	8193
Nijmegen	0.405 (0.493)	-0.172*** (0.059)	-0.157*** (0.060)		363
Tilburg	0.220 (0.415)	-0.040 (0.040)	-0.058 (0.039)		771

Notes: Estimates of ITT effects three months before the end of treatment. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating a modified version of Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes calendar quarter fixed effects at sites with multiple start cohorts and randomization strata fixed effects in Utrecht. *** p<0.01, ** p<0.05, * p<0.1.

Table 18. Employment Effects at Endline: Cut-offs Based on Earnings

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Earnings >70% of minimum wage (1/0)					
Utrecht	0.070 (0.255)	0.032 (0.029)	0.039 (0.028)	0.027 (0.027)	750
Deventer	0.115 (0.319)	-0.001 (0.022)	0.000 (0.023)		1547
Wageningen	0.163 (0.371)	0.014 (0.053)	0.049 (0.053)	-0.025 (0.048)	407
Groningen	0.097 (0.296)	-0.028 (0.024)	-0.022 (0.024)	-0.019 (0.024)	8190
Nijmegen	0.240 (0.429)	-0.070 (0.051)	-0.085* (0.051)		362
Tilburg	0.130 (0.337)	-0.014 (0.031)	0.006 (0.031)		771
Panel B: Earnings >30% of minimum wage (1/0)					
Utrecht	0.128 (0.335)	0.046 (0.033)	0.052 (0.033)	0.054* (0.033)	750
Deventer	0.184 (0.388)	-0.010 (0.025)	0.000 (0.026)		1547
Wageningen	0.293 (0.458)	0.013 (0.062)	0.018 (0.063)	-0.005 (0.061)	407
Groningen	0.145 (0.353)	-0.031 (0.026)	-0.025 (0.026)	-0.010 (0.026)	8190
Nijmegen	0.421 (0.496)	-0.158*** (0.054)	-0.159*** (0.056)		362
Tilburg	0.210 (0.408)	-0.013 (0.037)	-0.019 (0.036)		771

Notes: Estimates of ITT effects three months before the end of treatment. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, cumulative earnings in nine to six months before treatment, and outcome values in months 8, 16, and 24 before treatment. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 19. Employment Effects at Endline: Varying Cut-off Points

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Hours worked per week >15 (1/0)					
Utrecht	0.107 (0.310)	0.040 (0.032)	0.067** (0.033)	0.049 (0.032)	750
Deventer	0.166 (0.372)	-0.011 (0.024)	-0.016 (0.025)		1547
Wageningen	0.250 (0.435)	0.014 (0.061)	0.046 (0.062)	-0.034 (0.059)	407
Groningen	0.124 (0.331)	-0.028 (0.026)	-0.020 (0.026)	-0.008 (0.026)	8191
Nijmegen	0.372 (0.485)	-0.128** (0.055)	-0.134** (0.055)		362
Tilburg	0.200 (0.401)	-0.022 (0.036)	-0.039 (0.034)		771
Panel B: Hours worked per week >9 (1/0)					
Utrecht	0.123 (0.329)	0.046 (0.033)	0.070** (0.034)	0.080** (0.034)	750
Deventer	0.191 (0.394)	-0.005 (0.025)	-0.000 (0.026)		1547
Wageningen	0.293 (0.458)	0.032 (0.064)	0.034 (0.063)	-0.027 (0.062)	407
Groningen	0.159 (0.367)	-0.036 (0.026)	-0.025 (0.026)	-0.013 (0.026)	8191
Nijmegen	0.430 (0.497)	-0.143*** (0.055)	-0.119** (0.056)		362
Tilburg	0.220 (0.415)	-0.018 (0.037)	-0.025 (0.037)		771

Notes: Estimates of ITT effects three months before the end of treatment. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, cumulative earnings in nine to six months before treatment, and outcome values in months 8, 16, and 24 before treatment. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 20. Effects on Use of Prescription Drugs

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Use of prescription drugs (1/0)					
Utrecht	0.774 (0.419)	0.007 (0.036)	0.023 (0.036)	0.041 (0.036)	750
Deventer	0.807 (0.395)	-0.008 (0.026)	0.035 (0.025)		1561
Wageningen	0.761 (0.429)	0.047 (0.058)	-0.028 (0.057)	-0.003 (0.058)	407
Groningen	0.714 (0.452)	-0.031 (0.033)	-0.014 (0.033)	-0.014 (0.033)	8217
Nijmegen	0.711 (0.455)	0.017 (0.052)	0.055 (0.053)		362
Tilburg	0.810 (0.393)	-0.012 (0.034)	-0.038 (0.036)		775

Notes: Estimates of ITT effects in 2019. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, cumulative earnings in nine to six months before treatment, and outcome values in years 2014, 2015, and 2016. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 21. Health and Financial Effects: No Lagged Controls

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Expenditures GP visit (euro)					
Utrecht	227.491 (160.153)	-1.570 (16.614)	-11.185 (15.012)	-5.089 (14.899)	752
Deventer	223.842 (180.529)	-7.028 (11.831)	-2.369 (11.538)		1568
Wageningen	198.048 (153.855)	12.410 (20.704)	-9.974 (20.147)	-16.262 (20.205)	407
Groningen	219.062 (246.494)	-28.839 (20.795)	-23.725 (20.752)	-26.703 (20.763)	8236
Nijmegen	197.808 (147.519)	1.403 (18.012)	46.097 (33.191)		362
Tilburg	227.857 (153.560)	-0.867 (17.266)	-7.241 (15.653)		775
Panel B: Problematic debts (1/0)					
Utrecht	0.038 (0.191)	-0.004 (0.018)	-0.009 (0.018)	-0.008 (0.020)	748
Deventer	0.056 (0.230)	-0.002 (0.017)	-0.019 (0.015)		1547
Wageningen	0.033 (0.179)	0.007 (0.026)	-0.001 (0.025)	-0.011 (0.024)	407
Groningen	0.034 (0.183)	0.023 (0.016)	0.012 (0.016)	0.009 (0.016)	8162
Nijmegen	0.033 (0.180)	0.004 (0.025)	0.025 (0.029)		361
Tilburg	0.030 (0.171)	0.035* (0.021)	0.008 (0.016)		769

Notes: Estimates of ITT effects in 2019. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating a modified version of Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, and cumulative earnings in nine to six months before treatment. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 22. Health and Financial Effects: No Controls

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Expenditures GP visit (euro)					
Utrecht	227.491 (160.153)	-3.917 (16.730)	-8.096 (15.118)	-5.000 (15.094)	752
Deventer	223.842 (180.529)	-9.571 (11.865)	-4.676 (11.475)		1574
Wageningen	198.048 (153.855)	15.531 (21.954)	-12.052 (20.250)	-20.002 (19.160)	407
Groningen	219.062 (246.494)	-32.391 (20.751)	-27.932 (20.706)	-30.507 (20.810)	8238
Nijmegen	197.808 (147.519)	-0.342 (17.988)	44.630 (32.247)		363
Tilburg	227.857 (153.560)	8.244 (17.898)	-2.625 (15.376)		775
Panel B: Problematic debts (1/0)					
Utrecht	0.038 (0.191)	-0.006 (0.019)	-0.011 (0.019)	-0.011 (0.019)	748
Deventer	0.056 (0.230)	-0.002 (0.017)	-0.016 (0.015)		1553
Wageningen	0.033 (0.179)	-0.004 (0.025)	-0.001 (0.026)	-0.013 (0.023)	407
Groningen	0.034 (0.183)	0.022 (0.016)	0.010 (0.016)	0.008 (0.016)	8164
Nijmegen	0.033 (0.180)	0.009 (0.025)	0.034 (0.028)		362
Tilburg	0.030 (0.171)	0.024 (0.021)	-0.004 (0.017)		769

Notes: Estimates of ITT effects in 2019. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating a modified version of Eq.(1). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification include calendar quarter fixed effects for sites with multiple start cohorts and randomization strata fixed effects in Utrecht. *** p<0.01, ** p<0.05, * p<0.1.

Supporting Online Materials Text 3: Specification Local Average Treatment Effects

We estimate local average treatment effects by instrumenting actual treatment status with assigned treatment status in a two-stage least squares (2SLS) framework (Angrist & Imbens, 1994). To obtain conservative estimates, we count partially treated subjects as fully treated. The first-stage equation takes the following form:

$$\widehat{D}_{im} = \alpha^t + \sum_{m=1}^M \beta_m^t A_{im} + X_i' \gamma + \delta y_i^{t-x} + q + \varepsilon \quad (2)$$

and estimates actual treatment status, \widehat{D}_{im} , as a function of assigned treatment status, A_{im} , and the same variables that are included in Eq.(1). Note that Eq.(2) is a simplification of the system of first-stage equations, as we estimate a different first-stage for every treatment condition.

The second stage equation estimates treatment effects and takes the following form:

$$y_i^t = \alpha^t + \sum_{m=1}^M \theta_m^t \widehat{D}_{im} + X_i' \gamma + \delta y_i^{t-x} + q + \varepsilon \quad (3)$$

The parameters of interest are θ_m^t , which describe local average treatment effects, that is, the effects of treatment for the subgroup of compliers.

Tables S5: Local Average Treatment Effects

Table 23. Employment Effects at Endline: Local Average Treatment Effects

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Economic self-sufficiency (1/0)					
Utrecht	0.059 (0.236)	0.021 (0.029)	0.047 (0.031)	0.021 (0.028)	750
Deventer	0.087 (0.282)	0.035 (0.078)	0.037 (0.114)		1,547
Wageningen	0.109 (0.313)	0.077 (0.056)	0.093* (0.054)	0.010 (0.048)	407
Groningen	0.082 (0.274)				
Nijmegen	0.149 (0.357)	-0.038 (0.047)	-0.035 (0.051)		362
Tilburg	0.095 (0.294)	-0.007 (0.036)	-0.010 (0.037)		771
Panel B: Significant employment (1/0)					
Utrecht	0.118 (0.323)	0.045 (0.036)	0.078** (0.038)	0.078** (0.038)	750
Deventer	0.179 (0.383)	-0.036 (0.093)	-0.009 (0.137)		1,547
Wageningen	0.293 (0.458)	-0.003 (0.074)	0.039 (0.072)	-0.074 (0.066)	407
Groningen	0.144 (0.351)				
Nijmegen	0.405 (0.493)	-0.159*** (0.060)	-0.168** (0.066)		362
Tilburg	0.220 (0.415)	-0.029 (0.048)	-0.046 (0.049)		771
F-Statistics first stage					
Utrecht		541	386	370	
Deventer		44	29		
Wageningen		176	195	255	
Groningen					
Nijmegen		487	264		
Tilburg		102	214		

Notes: Estimates of local average treatment effects three months before the end of treatment. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating Eq.(3). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, cumulative earnings in nine to six months before treatment, and outcome values in months 8, 16, and 24 before treatment. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. The lower section reports *F*-statistics for the respective first stages. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 24. Health and Financial Effects: Local Average Treatment Effects

	Control mean (SD) (1)	Autonomy (2)	Counseling (3)	Earnings release (4)	N (5)
Panel A: Expenditures GP visit (euro)					
Utrecht	227.491 (160.153)	-5.946 (15.452)	-12.029 (15.388)	-13.873 (15.394)	752
Deventer	223.842 (180.529)	-32.822 (40.815)	-11.139 (55.615)		1,568
Wageningen	198.048 (153.855)	19.009 (22.238)	-12.996 (20.595)	-10.518 (20.488)	407
Groningen	189.665 (146.359)				
Nijmegen	197.808 (147.519)	11.122 (18.396)	62.073 (38.515)		362
Tilburg	227.857 (153.560)	8.300 (19.218)	-5.387 (18.149)		775
Panel B: Problematic debts (1/0)					
Utrecht	0.038 (0.191)	-0.013 (0.017)	-0.024 (0.017)	-0.012 (0.019)	748
Deventer	0.056 (0.230)	-0.013 (0.057)	-0.091 (0.078)		1,547
Wageningen	0.033 (0.179)	0.004 (0.029)	0.002 (0.029)	-0.018 (0.027)	407
Groningen	0.047 (0.211)				
Nijmegen	0.033 (0.180)	-0.010 (0.027)	0.026 (0.032)		361
Tilburg	0.030 (0.171)	0.042 (0.026)	0.005 (0.022)		769
F-Statistics first stage					
Utrecht		544	377	367	
Deventer		45	29		
Wageningen		160	199	252	
Groningen					
Nijmegen		488	262		
Tilburg		101	209		

Notes: Estimates of local average treatment effects in 2019. Column (1) reports control group means with standard deviations in parentheses. Column (2)-(4) show coefficients on the treatment dummies, estimating Eq.(3). Robust standard errors are in parentheses. The treatment is *autonomy* with *earnings release* and *counseling* with *earnings release* in Deventer and Nijmegen, and *autonomy* with *earnings release* and *work bonus* and *counseling* with *earnings release* and *work bonus* in Tilburg. The reference group is the default control group for all sites except Groningen, where the reference group is a randomly selected second control group. The specification includes control variables for age, gender, education level, migration background, household type, duration of welfare spell, cumulative earnings in nine to six months before treatment, and outcome values in 2014-2016. For sites with multiple start cohorts, calendar quarter fixed effects are added. The specification for Utrecht includes randomization strata fixed effects. The lower section reports *F*-statistics for the respective first stages. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.