UNIVERSITEIT AMSTERDAM

VU Research Portal

Caseworker's discretion and the effectiveness of welfare-to-work programs

Bolhaar, Jonneke; Ketel, Nadine; van der Klaauw, Bas

Link to publication in VU Research Portal

Caseworker's discretion and the effectiveness of welfare-to-work programs ☆

Jonneke Bolhaar [a], Nadine Ketel [b], Bas van der Klaauw [c],*

[a] Netherlands Bureau for Economic Policy Analysis (CPB), Department of economics, VU Amsterdam, De Boelelaan 1105, Amsterdam, 1081 HV, The Netherlands
[b] VU Amsterdam, University of Gothenburg, IZA, CEPR, Department of economics, VU Amsterdam, De Boelelaan 1105, Amsterdam, 1081 HV, The Netherlands
[c] VU Amsterdam, Tinbergen Institute, Department of economics, VU Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands

ARTICLE INFO

ABSTRACT

In this paper we focus on the role of caseworkers in the assignment and take-up of welfare-to-work programs. We conduct a field experiment that generates exogenous variation in the assignment of caseworkers to different policy regimes. The experiment allows us to provide evidence on the effectiveness of welfare-to-work programs and to study how caseworkers exploit their discretion in assigning these programs to welfare recipients. We find substantial heterogeneity in how caseworkers assign welfare-to-work programs. Participation in the experiment and learning about the effectiveness of the different programs do not induce caseworkers to focus more on the effective programs. Obtaining knowledge about welfare-to-work programs is thus not enough to improve policy, also effort on implementation is required.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Randomized control trials (RCTs) are increasingly used to evaluate the effectiveness of active labor market programs. In their 2010 meta analysis on active labor market programs, Card et al. (2010) report that 10% of the studies use an RCT explicitly designed to empirically evaluate a program. In the 2018 update this increased to 19% (Card et al., 2018). RCTs solve the problem of selective participation in active labor market programs and, therefore, are considered to provide a credible empirical evaluation (e.g. Heckman et al., 1999). However, using RCTs some problems remain unsolved (Rothstein and Von Wachter, 2017). For example, the estimated treatment effect is only policy relevant if program participation can easily be varied, which requires that caseworkers have limited discretionary power when deciding about participation in active labor market programs.

In this paper, we focus on the role of the caseworker in the assignment and take-up of active labor market programs. RCTs assign all unemployed workers in the treatment group to participate in the program, but usually ignore the take-up decision. The strict program assignment rule of an RCT does not concur with the large degree of discretion which caseworkers have in many countries (e.g. Behncke et al., 2009, Bell and Orr, 2002, Lechner and Smith, 2007, Schmieder and Trenkle, 2016, Vikström, 2017).[1] The average treatment effect estimated using the RCT may then not be the most policy relevant treatment effect. While there is quite a lot of recent evidence on the effectiveness of various active labor market programs, for example documented in Card et al. (2018), much less is known about

[1] Schiprowski (2020) and Rosholm (2014) find that meetings with the caseworker are important and that there is substantial heterogeneity between caseworkers. Vikström (2017) refers to a PhD thesis by Eriksson (1997), which shows that caseworker heterogeneity is more important than heterogeneity among unemployed workers when assigning active labor market programs. Similar results are reported by Huber et al. (2017).

the motivations and goals of caseworkers (Schmieder and Trenkle, 2016).

This paper contributes to the literature by answering three questions. First, we investigate the effectiveness of a set of existing active labor market programs to support welfare benefits recipients and study heterogeneity in treatment effects on a series of labor market outcomes. Second, we study how the use of these active labor market programs varies between locations and caseworkers. In our setting caseworkers have a limited caseload and meetings with benefits recipients occur frequently. Caseworkers have an excellent opportunity to gain detailed information about the benefits recipient's ability and needs, which would allow them to effectively match benefits recipients to services (Lechner and Smith, 2007). We relate the use of active labor market programs to heterogeneity in effects of participation in these programs. In addition, we explore if caseworkers prefer certain programs, which they offer to most of their benefits recipients and if caseworkers which express such specialization in a particular program obtain better results when providing this program to a benefits recipient. Third, we study knowledge of caseworkers by investigating if caseworkers change their preferences for programs after having been exposed to a different (and possibly more effective) way of working and having learned about the effectiveness of all programs.

We conduct a field experiment covering all new entrants into welfare benefits (with a potential to work) in Amsterdam in the period April 2012 to March 2013.[2] The field experiment generates exogenous variation in the assignment to three different welfare-to-work programs: direct job matching, job-search training and counseling. We evaluate these programs against two alternatives, one where the caseworker has full discretion in choosing the program which she finds most appropriate and one where the benefits recipient does not participate in any program. This setup allows us to evaluate each of the programs against the alternative of no program and to study what the added value of caseworkers is in selecting an effective program. In addition, the program choices when the caseworker has full discretion identify specialization of the caseworker. We address this specialization and the beliefs about the effectiveness of the different programs in a survey among caseworkers which provides insights in learning by caseworkers after having gained experience with other programs.

Our study relates to a substantial literature evaluating active labor market programs summarized in Card et al. (2018). Most RCTs focus on a program additional to the standard support provided to benefits recipients (e.g. Graversen and Van Ours, 2008, Van den Berg and Van der Klaauw, 2006). A unique feature of our field experiment is that we introduce a treatment where no support is given, which allows us to evaluate the existing programs rather than only new programs. Our study also relates to the literature on the optimal assignment of unemployed workers to active labor market programs. Lechner and Smith (2007) compare caseworker discretion, a statistical treatment rule based on observable participant characteristics and random assignment to services and find that caseworkers obtain roughly the same post-program employment rate as the random allocation, while statistical treatment rules outperform both. Compared to our study, they measure program effectiveness by only controlling for observed participant characteristics rather than using exogenous variation from a field experiment. Behncke et al. (2009) focus on targeting active labor market programs using a large field experiment in Switzerland. They find that caseworkers ignore the information of the statistical system which is provided to them and

thus do not change behavior. However, use of the statistical system was neither encouraged nor monitored. Also, Schmieder and Trenkle (2016) find that caseworkers do not optimize their behavior when the length of the entitlement period to unemployment insurance benefits changes.

For the empirical analysis we combine data from different sources to construct a very detailed administrative dataset describing the participants in our field experiment. We observe individual labor market histories in the years prior to collecting welfare benefits and have measures of guidance received at the welfare agency such as meetings with caseworkers and participation in welfare-to-work programs. After exit from the welfare benefits we observe employment status and earnings at the weekly level. Our empirical results show that the effects of the various welfare-to-work programs differ substantially. Whereas direct matching improves the labor market outcomes of benefits recipients, participation in job-search activation has mainly adverse effects. We find only limited evidence for heterogeneity in the effects of participating in a program and this heterogeneity only relates to characteristics of the welfare recipient and not to the caseworker. In particular, caseworkers who use a particular program very often when having full discretion obtain the same average effects of this program as caseworker who only rarely use the same program. Also, we see that local offices adopt different approaches under full discretion while we find little support for heterogeneous treatment effects. From this we conclude that not all caseworkers use their discretion optimally in assigning benefits recipients to programs. This is confirmed by results from our survey among caseworkers which show that caseworkers do not easily change their preferences for welfare-to-work programs. The policy implication is that even if a randomized controlled trial provides credible evidence on the effectiveness of certain treatments, the roll-out of such policies is not obvious.

The remainder of this paper is structured as follows. The next section provides details about the welfare system in the Netherlands, the specific setting of the experiment and describes the role of caseworkers. In Section 3 we explain the experimental design of the field experiment. Section 4 describes the data and provides evidence on the randomization and compliance rates. In Section 5 we present the main results, and Section 6 concludes.

2. Institutional background and the role of caseworkers

2.1. Welfare benefits in the Netherlands

Welfare provides benefits to households that do not have enough income. The benefits level depends on the composition of the household.[3] Monthly benefits range from 668 euro for a single without children, to 1336 euro for a couple with children. In the same period the net minimum wage was about 1200 euro per month. Individuals with a (part-time) job earning less than the welfare benefits level, can receive partial welfare benefits, which complements their income from work to the welfare benefits level. In that case, the marginal tax rate for (additional) earnings is 100 %. There is no limit to the period that households can collect welfare benefits.

Welfare recipients have the obligation to accept any type of employment, also if it does not fit their education or work experience. The rules on eligibility for welfare and the level of the benefits are determined at the national level, but municipalities decide about welfare-to-work programs and other activation policies for welfare recipients. Individuals have to apply for welfare benefits in the municipality where they live.

---

[2] This field experiment was implemented simultaneously with the experiment evaluating job-search periods prior to entry in welfare benefits which is described in Bolhaar et al. (2019). The random assignment in both field experiments is orthogonal, which allows to evaluate them separately.

[3] Table A1 in the Appendix gives an overview of the different benefits levels by household composition in the year 2012 (start of our observation period).

	Mean	N
Female	60%	50
Age	41.9	50
Bachelor/master degree	96%	50
Years worked at welfare agency	10.6	50
Being caseworker is first job	23%	44
Ever received benefits	29%	49

Note: The information in this table is based on a survey among the caseworkers that participated in the experiment. The response rate to the survey is 81%.

2.2. Caseworkers

Welfare applicants in Amsterdam are randomly matched to a caseworker (within the welfare office), which allows the benefits agency to benchmark caseworkers against each other.[4] Applicants are supposed to meet their caseworker twice a month. These meetings with the caseworker are not necessarily face-to-face meetings but can also be phone calls or email contacts. The caseworker supports the welfare recipient in her job search, can offer participation in welfare-to-work programs, monitors job-search effort and sanctions individuals that do not comply with the job-search requirements.[5] The welfare-to-work programs are discussed in more detail in the next section.

During the first meeting the caseworker determines the labor market prospects of the welfare applicant using a computerized program that profiles based on characteristics such as work history, age, education, language and computer skills, family situation and physical or psychological problems. This results in a classification into four classes that determines which type of guidance a welfare benefits recipient receives and what the obligations of the welfare recipient are. We focus on class IV, which contains welfare recipients who do not have work limitations and are considered to be able to find a job within six months. Welfare recipients in this class have to actively search for work (typically they are required to make one or two job applications each week). About 40% of all welfare applicants in Amsterdam are assigned to class IV.[6] Welfare offices have teams of caseworkers that support only welfare recipients in class IV in their job search. If the welfare recipient does not find work within six months, she should be transferred to class III and another team of caseworkers. In practice, the period in class IV is often extended with some months.

Table 1 provides characteristics of the caseworkers for the welfare recipients in class IV. These caseworkers are on average 42 years old, almost all hold a bachelor and/or master degree and 60% are female. Average tenure at the welfare agency is long (almost 11 years), although for only 23% of the caseworkers it is their first job. Finally, 29% of the caseworkers have ever received either welfare or unemployment insurance benefits.

Fig. 1. Self-assessed heterogeneity in program choice of caseworkers. Note: data from post-experiment survey among caseworkers (response rate 81%).

2.3. Welfare-to-work programs and caseworker discretion

A caseworker assigned to class IV can use three programs to support welfare recipients in their job search. First, they can send the individual to a job-search activation course. This is an intensive eight-week program with daily sessions. During the first two weeks welfare recipients receive job-application training and guidance on how to find vacancies. During the subsequent six weeks participants spend a few hours per day in a computer room, where they make job applications (under the guidance of a trainer). Second, the caseworker can match the individual directly to a vacancy, which is taken from a pool of vacancies gathered by a separate unit in the welfare office. The match is often accompanied by a trial period or a wage subsidy for the employer.[7] Third, the caseworker can apply caseworker counseling. This includes regular meetings (typically every other week) in which the caseworker helps the individual with the job-search process, setting up the C.V., etc. Before our experiment, the caseworker had full discretion in the assignment of programs to welfare recipients with the restrictions that direct matching to vacancies can only be done in case of a suitable vacancy and the eight-week job-search activation course has entry requirements on language and computer skills.

The welfare-to-work programs offered by the welfare administration of Amsterdam are very similar to programs used in many other municipalities, and also, in other countries. Data from Eurostat shows that most European countries spend substantial amounts on what they call client services and employment incentives. The job-search activation course and caseworker counseling fall in the category of client services. According to Eurostat expenditures on client services are highest in Denmark, France, UK, Germany and Sweden. Direct matching contains most elements considered in employment incentives such as temporary financial incentives to employers. Card et al. (2010) show that in Nordic countries, but also in Anglo-Saxon countries a very substantial share of the evaluation studies concern subsidized private sector employment, such as direct matching. Many of the programs that are evaluated in this literature have strong similarities with the programs considered in this paper.

In a post-experiment survey, we asked the caseworkers to indicate how much they target programs to welfare recipients. On a zero to one hundred scale caseworkers answer how much they agree with two statements (see Fig. 1 for a summary of the answers). The first question concerned heterogeneity among welfare recipients. Almost all caseworkers agree that heterogeneity in individual characteristics is important when assisting welfare recipients. The second question concerned the choice for welfare-to-work programs. The caseworkers indicate that they do not apply a uniform policy to support welfare recipients in their job search.

To investigate further how caseworkers target welfare-to-work programs to welfare recipients, we asked the caseworkers how they

---

[4] Note that we cannot use the average treatment assignment rate of a caseworker as an instrumental variable to estimate the effect of the treatment, as Maestas et al. (2013) and Dahl et al. (2014) do. Caseworkers usually interact more with their clients than judges and may also provide guidance unrelated to the decided treatment. If this type of guidance is related to how frequent caseworkers choose certain active labor market programs, the validity of an instrumental variable approach is violated.

[5] A sanction generally reduces benefits with 30% for one month.

[6] Individuals in the three other classes have a larger distance to the labor market. Class III individuals should be able to work, but lack some (social) skills and require guidance to find work and stay employed. These individuals often start working in a subsidized job. Individuals in class I and II have social problems and/or physical limitations that make them unfit for work.
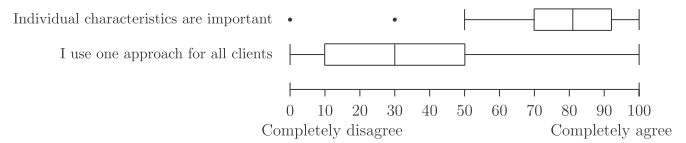
---

[7] A trial period is a period of one to three months in which a welfare benefits recipient works for an employer but receives welfare benefits instead of a wage. Trial periods are aimed at employers that have hesitations about the capability of welfare benefits recipients and are meant as a possibility for welfare benefits recipients to prove that they are suitable for the job.
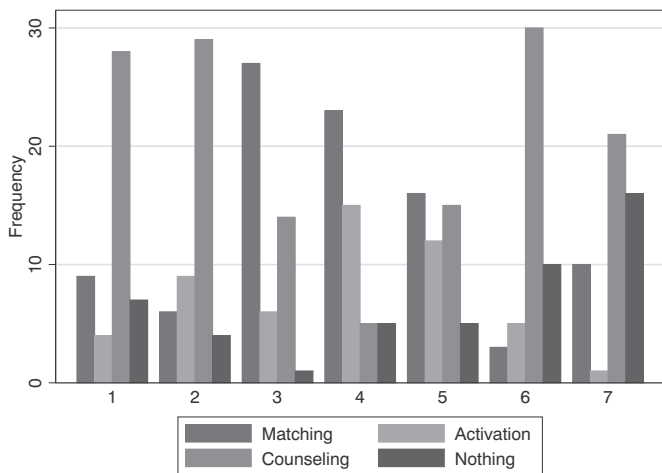
Fig. 2. Frequency of program choices of caseworkers to fictional welfare recipients. Note: Data from post-experiment survey among caseworkers (response rate 81%). Caseworkers are asked to state which welfare-to-work programs they use for seven fictional welfare recipients (numbers on the x-axis), which are described in Table A2 in the Appendix.

would support a number of hypothetical welfare recipients.[8] For example, client number one is 'a married man of 50 years old with limited command of Dutch and only a primary school degree'. Fig. 2 shows how many caseworkers would use each of the welfare-to-work programs to support each of the seven individuals. Three things are important to note. First, the program which is chosen most often varies between the seven cases. Second, caseworkers almost always prefer to provide a welfare-to-work program over having a passive role and giving the welfare recipient time to search for work herself. And third, in all cases a substantial number of caseworkers deviates from the most often chosen welfare-to-work program. The results confirm that heterogeneity among welfare applicants is taken into account when offering welfare-to-work programs and show that caseworkers have different opinions about the most suitable welfare-to-work program and have a lot of discretion in providing these programs.

3. The experiment

3.1. Experimental design

The field experiment includes five policy regimes.[9] For each of the three welfare-to-work programs, there is a policy regime in which caseworkers are supposed to assign as many welfare benefits recipients as possible to the specific welfare-to-work program. In the fourth policy regime, benefits recipients do not receive any guidance from the caseworker and are not supposed to participate in any welfare-to-work program. This is the control group, which allows us to evaluate the effectiveness of the three welfare-to-work programs. In the fifth policy regime the caseworker has the usual (full) discretion in assigning benefits recipients to the welfare-to-work programs (or to provide no guidance). With this policy regime we

can test whether caseworkers choose the most effective approach for benefits recipients.

During the experiment, caseworkers are assigned to one of the five policy regimes. They are asked to treat all new clients they receive according to the rules of this policy regime for at least six months. If following the policy regime is inappropriate or not suitable for the client, caseworkers can 'opt out', and use another program to guide a client back to work. We prefer to randomize policy regimes rather than participation in a particular welfare-to-work program. In reality caseworkers always have some discretion or they will not be able to convince some benefits recipients to participate. Therefore, policy makers at the welfare agency do not use strict protocols which impose full compliance, but rather formulate policy regimes with some discretion for caseworkers. A limitation of our experimental setting is that in reality policy makers can fully close down certain welfare-to-work programs, while in our setting all programs remain existing. The latter makes it easier for caseworkers to deviate from the rules in a restricted policy regime.

Every three months the policy regime of the caseworker for *new* welfare benefits recipients changes. This implies that caseworkers support benefits recipients in different policy regimes at the same time, as they have to treat an individual according to the same policy regime for at least six months.[10] Welfare recipients apply at five different locations in Amsterdam, and each location has their own team of caseworkers. We made sure that in every three-month period each of the five policy regimes is allocated to at least one member of each team. At the start of a new three-month period, each caseworker was instructed by us about her new policy regime.

The five policy regimes are communicated to the caseworkers as follows:

Counseling Provide counseling to the welfare recipients. Do *not* act as an intermediary between vacancies and the welfare recipient and do *not* offer participation in the job-search activation program.

Direct matching Try to match the welfare recipient directly to a vacancy. You can offer a trial period or wage subsidy to an employer if that helps to establish a match. If there are no appropriate vacancies, provide counseling in the meantime. Do *not* offer participation in the job-search activation program.

Job-search activation Offer participation in the job-search activation program to the welfare recipient. Provide counseling if the welfare recipient lacks the (computer or language) skills to participate, is on the waiting list or has completed the program. Do *not* act as an intermediary between vacancies and the welfare recipient.

Nothing Do not use any welfare-to-work program. Do not initiate contact with the welfare recipient but be available for questions of the welfare recipient.

Full discretion Choose the program(s) that you think are most appropriate. So follow your usual approach of supporting the welfare recipient.

The policy regimes are based on the programs that the welfare agency offered before our experiment. The caseworkers were thus experienced in the use of each program. Comparing the policy regime for each of the three programs with the policy regime with no guidance provides insight in the effectiveness of the three welfare-to-work programs. If we compare the policy regime with

---

[8] Table A2 in the Appendix provides the details of the hypothetical welfare recipients as they were presented to the caseworkers. These welfare recipients closely resemble actual applicants for welfare benefits at the time of our experiment.

[9] The original design of the experiment, including a power analysis, can be found at http://personal.vu.nl/b.vander.klaauw/ResearchProposalDWI2012.pdf.

[10] This feature makes the design complicated for the caseworker. To remind them of their current policy regime, we handed out forms with the policy regime preprinted. Caseworkers were asked to fill in these forms at the intake meeting. Most caseworkers also note the policy regime in the digital file of each individual, so that they remember which policy regime the individual is in.

full discretion for the caseworker with no guidance, we can assess the added value of the combination of all services provided by the welfare agency. To obtain insight in how well caseworkers allocate programs to benefits recipients, we can compare the policy regime of full discretion to the three policy regimes for each of the welfare-to-work programs.

The five policy regimes are randomly allocated to caseworkers and apply to all new welfare benefits recipients in a period of three months. New welfare recipients are within a local office randomly assigned to caseworkers.[11] The number of caseworkers assigned to the policy regimes is such that overall the number of participants in the job-search activation program and in direct matching remains similar as before the experiment. The policy regimes only apply to new welfare recipients, and not to the existing caseload. Furthermore, caseworkers have welfare recipients who are treated under different policy regimes at the same time. Roughly, the amount of time that caseworkers have available for each individual in their caseload will not vary much.[12]

Our setup had several advantages. Full randomization in the context of welfare-to-work programs is often difficult to enforce and raises ethical concerns. This is especially the case when evaluating existing policies (in contrast to the evaluation of additional policies or resources) as we achieve when introducing the policy regime nothing. The possibility to opt out in each policy regime gives caseworkers some discretion to prevent harmful effects of the experiment. This limited discretion also increased the support among caseworkers to commit to the experiment. The risk is of course that compliance to the policy regimes is low. We monitored this throughout the experiment by checking data on participation in programs. A second risk of our setup is that caseworkers exchange the welfare recipients they support. We monitor this by checking if welfare recipients meet another caseworker after the intake meeting and by checking the forms filled in at the intake meeting. Furthermore, in the next section we show balancing of the welfare recipients in the different policy regimes.

3.2. Implementation

The experiment was conducted in Amsterdam from April 2012 until September 2013. On January 1, 2012, Amsterdam had 790,110 inhabitants. During the time of our experiment there was a recession in the Netherlands and unemployment was relatively high. Within the population between 20 and 65 years old in Amsterdam, 6.4% received welfare benefits, compared to 3.1% for the whole of the Netherlands. Inflow into welfare benefits exceeded outflow, but at the same time the pool of applicants for welfare benefits had, on average, relatively more favorable characteristics than in better economic times.

The sample consists of all new welfare recipients older than 27 years that started collecting benefits between April 2012 and March 2013 and are classified as class IV (able to find a job within six months). The welfare agency in Amsterdam has five local offices in which each serves different neighborhoods of the city. Benefits recipients are not informed about their participation in an experiment, but the setup of the experiment was discussed with the formal council of welfare recipients.

Before the start of the experiment we organized meetings at all the local offices to inform the caseworkers about their role in the experiment. During the experiment we visited all local offices almost weekly to answer questions of caseworkers and monitor the implementation of the experiment.[13] At the start of every three-month period we instructed each caseworker individually about the new policy regime assigned to her. The caseworkers then also received a new set of forms which they were required to fill in for each new welfare recipient at the intake meeting. These forms were personalized for each caseworker and had their current policy regime pre-printed on the form, in order to remind the caseworker of their policy regime.

At one local office the manager was changed several times during the experiment. As a result, the caseworkers at this office received mixed instructions with respect to their participation in the experiment, which affected their compliance with the experiment. For example, the second (interim) manager explicitly instructed caseworkers to ignore the experiment to boost exit to work for the period that she was manager. We exclude all welfare recipients from this welfare office from the empirical analysis.[14]

After the experiment we administered a survey among all participating caseworkers. At the moment of completing the survey, the caseworkers were already informed about the main findings of the experiment via presentations at each local office (see Table A3 in the Appendix). The goal of this survey was to learn about choices caseworkers make when supporting benefits recipients in their job search. Furthermore, we wanted to have insight in the beliefs of caseworkers about the effectiveness of the different welfare-to-work programs and how they update these beliefs. This allows us to address caseworker learning.

4. Data

4.1. Data sources

In the empirical analysis we employ data from several sources. First, we use administrative data from the welfare agency of Amsterdam, with information on the start and end date of collecting welfare benefits, exact records of all benefits payments, and an identifier for the caseworker assigned to the benefits recipient. In addition, these data include all contacts (meetings, phone calls and email contact) between the caseworker and the benefits recipient and participation in the welfare-to-work programs. The welfare agency also registers individual characteristics such as gender, date of birth, highest obtained education and household composition.

The second data source is employment and income data abstracted from social insurance records. These records contain weekly information for each individual on the amount of earnings from employment and all types of benefits payments (including welfare benefits in other municipalities and benefits from other schemes such as unemployment and disability insurance). The data on all participants in the experiment cover the period from January 2008 until May 2014 and allow us to construct outcomes such as earnings from work and total income after random assignment. We exploit the retrospective nature to construct control variables describing the labor market history before entering welfare. Income from self employment is missing because self-employed workers do not participate in social insurance schemes.

We merge the data from these two sources to data available at Statistics Netherlands to follow individuals beyond one year after they have applied for welfare benefits. However, the data from Statistics Netherlands are less detailed than the information we have from the social insurance records. Specifically, amounts of benefits payments and earnings are less precise and aggregated over longer time

---

[11] At each local office, new welfare recipients are assigned to the caseworker with the lowest caseload. Such random assignment allows the welfare agency to evaluate caseworkers based on their (unconditional) realized outflow.
[12] At the start of the experiment the managers promised to adjust individual caseworker targets to the policy regime assigned to them.

[13] Table A3 in the Appendix shows a summary of all visits that we made to the local offices of the welfare agency during the experiment.
[14] Applicants from this local office constitute 20% of the experimental sample.

Table 2
Descriptive statistics of welfare recipients and balancing.

| | Policy regime | | | | | |
	Counseling (1)	Activation (2)	Matching (3)	Discretion (4)	Nothing (5)	All (6)
Female	36%	38%	38%	36%	38%	37%
x Partner	9%	13%	11%	10%	13%	11%
Children	10%	14%	13%	13%	10%	12%
Age (in years)	38.9	38.8	38.3	37.2**	38.9	38.4
Education (in years)	12.4	12.3	12.2	12.1	12.1	12.2
Education missing	1%	2%	2%	1%	1%	2%
Annual income 2 years before (×1000 €)	14.5	13.8	14.1	12.8	13.4	13.7
N	353	333	328	394	271	1679

Note: Stars indicate that there is a significant difference with the policy regime nothing. These p-values are weighted by the office of registration, as randomization took place within the welfare office. *** =significant at 1% level ** =at 5% level, * =at 10% level.

periods. Therefore, we only use these data when we look at outcomes in the long run (three years after applying for welfare benefits).

The forms completed by caseworkers at the intake meeting are the fourth data source. The form asks for the date of birth, gender, educational level, household composition, reason of applying for welfare, a subjective measure for the financial situation, an estimate from the caseworker on the duration of welfare benefits dependency and an indication which welfare-to-work programs will be offered to the welfare recipient. This final question was included to check whether the caseworker complied to the policy regime or used the possibility to deviate in special cases. The forms were filled in for 73% of the welfare recipients that participated in the experiment. Given that all crucial information is available through the administrative records (for the full sample), we only use these forms for complementary information.

Finally, we have information from the (ex-post) survey that was administered among the participating caseworkers. The survey includes questions on caseworker characteristics and asks their opinion on the experiment. This survey was conducted after the results of the evaluation were presented to the management of the welfare agency and at all local offices. When filling in the survey the caseworkers were familiar with the estimated effectiveness of each of the welfare-to-work programs. In the survey we asked for the caseworker's beliefs on effectiveness of the different programs and whether they changed their beliefs after the experiment. The response rate to the survey is 81 %(50 out of 62 caseworkers that participated in the experiment), or excluding the local office that did not comply to the experiment 78 %(40 out of 51 caseworkers). Weighted by the number of applicants per caseworker, the response rate is 87 %(83% excluding the non-complying welfare office), indicating that caseworkers that assisted more welfare recipients are more likely to complete the survey.

4.2. Descriptive statistics and balancing

During the experiment 2103 individuals that started collecting welfare benefits in Amsterdam satisfied the criteria for participation in our experiment (class IV and older than 27 years). A small number of individuals experienced multiple welfare spells within the experimental period, so the sample contains 2061 unique individuals. This number is slightly lower than the inflow of 2500 individuals which was expected prior to the experiment. From our sample we exclude 424 applications (416 unique individuals) from the local office where compliance was low due to changing managers during the experiment. Our final sample consists of 1679 spells of collecting welfare benefits experienced by 1645 unique individuals.

Table 2 provides descriptive statistics for the full sample (final column) and for each policy regime (the first five columns). The policy regime with full discretion for the caseworker has the largest

number of observations, while nothing has the fewest. In our design we ensured that all caseworkers are observed at least once in the policy regime with full discretion and we minimized the number of welfare recipients without any guidance, following the request of the welfare agency. The compositions of the five treatment groups are well balanced. Out of 28 t-tests, there is only one significant difference (at the five percent level) between the groups.[15]

Less than 40% of the welfare recipients are women, about 90 percent are single and slightly over 10% have children. The average age at starting collecting welfare benefits is 38.4 years. This is relatively young given that we only consider individuals older than 27. The explanation is that the entitlement period to unemployment insurance benefits depends on age, so older workers who lose their job have more time to find work before becoming dependent on welfare benefits. On average, the individuals have slightly more than 12 years of education. About 30% either have a bachelor or master degree and 35% completed higher vocational education. Slightly less than 15% have only followed primary education. In the two years before entering welfare, the average annual income was about 13,700 euro, which is approximately the minimum wage.

5. Results

We present results in four steps. First, we provide evidence on the effectiveness of the different policy regimes. Second, we investigate the choices that caseworkers make within each policy regime and use instrumental variables to estimate the effect of participating in each welfare-to-work program. Third, we investigate heterogenous treatment effects and study if caseworkers that specialize in a specific program get better results when applying this program. Finally, we investigate if caseworkers learn by assessing whether they focus on programs differently after being exposed to a different way of working and learning about the effectiveness of the programs.

5.1. Effects of the policy regimes

Fig. 3 shows for each policy regime the percentage of individuals that collect welfare benefits since their moment of application.[16] During the first 10 weeks after application, there are no substantial

---

[15] The p-values are weighted by office of registration, as randomization took place within each local office.

[16] The moment of application may be earlier than the moment of eligibility for welfare benefits. For example, the unemployment insurance administration advises individuals who are close to exhausting unemployment insurance benefits to already apply for welfare benefits. Also, some applications are initially incomplete and denied, but after a reapplication the individuals may still entitle to welfare. At the initial application, an individual is matched to a caseworker, which determines the policy regime.
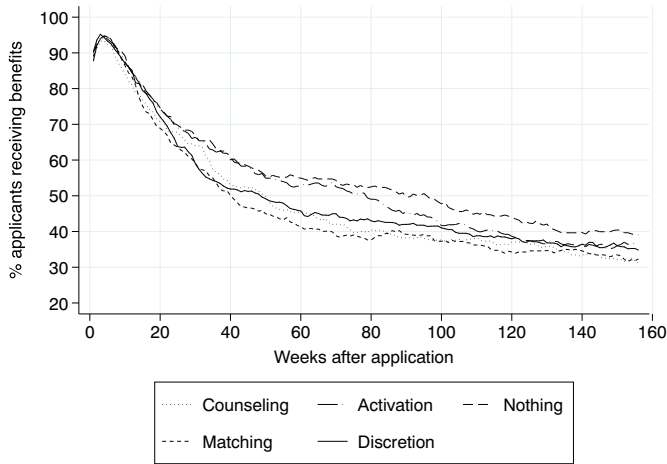
Fig. 3. Percentage of applicants receiving welfare benefits by policy regime. Note: Data from post-experiment survey among caseworkers (response rate 81%). Caseworkers are asked to state which welfare-to-work programs they use for seven fictional welfare recipients (numbers on the x-axis), which are described in Table A2 in the Appendix.

differences between policy regimes in the likelihood to receive benefits. After that exit in the direct matching regime is slightly higher than in the other policy regimes. After about 20 weeks differences between policy regimes get more pronounced. From then on, outflow in the policy regimes with no support and job-search activation lags behind. Between 40 and 80 weeks differences are largest, up to almost 15 percentage points. Outflow is highest in the policy regime with direct matching and lowest in the regime with no support. After 80 weeks the average benefits receipt in the different policy regimes converges somewhat. However, there is still an almost 10 percentage points difference between direct matching and no support three years after application.

To explore whether the observed differences in Fig. 3 are statistically significant, we estimate the following regression model:

$$Y_{i\tau tw} = \alpha_{\tau t} + \gamma_{wt} + \delta_t^c C_{i\tau w} + \delta_t^a A_{i\tau w} + \delta_t^m M_{i\tau w} + \delta_t^d D_{i\tau w} + X_i \beta_t + U_{i\tau tw} \quad (1)$$

We focus on estimating the effects of the different policy regimes on various labor market outcomes. As reference we take the policy regime with no guidance. The variable $Y_{i\tau tw}$ denotes the labor market outcome of individual i observed t time periods after applying for welfare benefits at welfare office w at calendar time τ. The four different policy regimes are indicated with C (counseling), A (activation), M (direct matching) and D (full discretion of the caseworker on which welfare-to-work program to apply). The vector X_i contains background characteristics including age, gender, partner status, having children, cumulative income in the 24 months before applying to welfare and level of education. In addition, we control for whether the applicant received a job-search period at the time of application for welfare benefits (for a discussion of this policy, see Bolhaar et al. (2019)).[17] The parameters $\alpha_{\tau t}$ are fixed effects for the quarter of entering welfare and account for business cycle effects.

The parameters γ_{wt} are the fixed effects for the local welfare offices, which control for potential differences between the local labor markets in the five city districts. Furthermore, recall that within local offices the policy regimes are randomly assigned to caseworkers and new welfare recipients are randomly allocated to caseworkers. The fixed effects for the local offices are crucial to deal with the conditional random assignment (i.e. the rates at which

policy regimes are assigned differ between local offices). In the previous subsection we showed that characteristics of welfare recipients in the different treatment groups are balanced after weighting for the local office.[18] We estimate this model separately for different elapsed durations t since applying for welfare benefits. The parameters of interest $\delta_t^\bullet$ describe the effects of the different policy regimes t weeks after entering welfare benefits. Using this empirical model, we obtain the intention-to-treat effects (compared to the regime with no support of the caseworker).

Fig. 4 presents the estimation results of Eq. (1) (together with the 90% confidence interval) where benefits receipt is the outcome. Outflow in the policy regime with job-search activation is the same as in the policy regime without support. The policy regime with counseling has a higher outflow after 30 weeks, but differences are not significant. Differences are significant for direct matching and for full discretion of the caseworker. These two policy regimes show very similar patterns of outflow.[19]

Finding work is the most frequent reason for outflow from welfare benefits, but it is not the only reason. Furthermore, individuals who leave welfare because they find work, may lose their job again. In Fig. 5 we show effects on having work (with positive earnings) for each week after application to welfare benefits. The figures largely show the reversed pattern as in Fig. 4. The main exception is that after about 40 weeks individuals in the job-search activation regime have a significantly *lower* probability to have a job with positive earnings than in the policy regime with no support.

Welfare-to-work programs may not only affect the rate at which work is found but can also affect the quality of the job. Job quality is often proxied by earnings. Fig. 6 reports the effects of the policy regimes on earnings in each week after application. Individuals without a job in a particular week, have zero earnings. There are also individuals with very flexible contracts who have some weeks with very low earnings. Since the earnings measure is noisier than, for example, an indicator for work, confidence intervals are wider. The effects of direct matching and full discretion are no longer significant, which suggests that many people find low paying jobs. This is in agreement with the observation that trial periods and wage subsidies are often necessary to convince employers to hire welfare recipients.

Most striking is the negative effect of the job-search activation regime on earnings, which already becomes negative after about 13 weeks. The job-search activation program was originally designed for workers in lower classes (with a larger distance to the labor market). The focus is largely on finding employment via temp work agencies. Because there were not enough welfare recipients in the lower classes satisfying the participation criteria, the job-search activation was also made available for welfare recipients in class IV. Our results suggest that jobs in the temp work sector are not the best match for welfare recipients in class IV.

Finally, Fig. 7 shows the effects on total income, which is the sum of welfare benefits, earnings and income from other types of benefits. This figure confirms the earlier findings. The higher job finding in the regimes with direct matching and full discretion does not translate in higher income for the welfare recipient. The earnings in most jobs do not exceed the benefits level. Job-search activation significantly reduces total income. The most likely explanation is that individuals who exit the welfare benefits system get flexible contracts in the temp work sector which often provide too few hours of work in a week to earn more than the welfare benefits level. Counseling has zero effects on total income.

---

[17] Controlling for a job-search period does not affect our results.

[18] Since a small fraction of the applicants appear twice in the data we cluster standard errors at the level of the individual welfare applicant.
[19] We have also looked at the effects of the policy regimes on the amount of welfare benefits payments, to take both full and partial benefits receipt into account. The results for this outcome are very similar to the results for any benefit receipt.
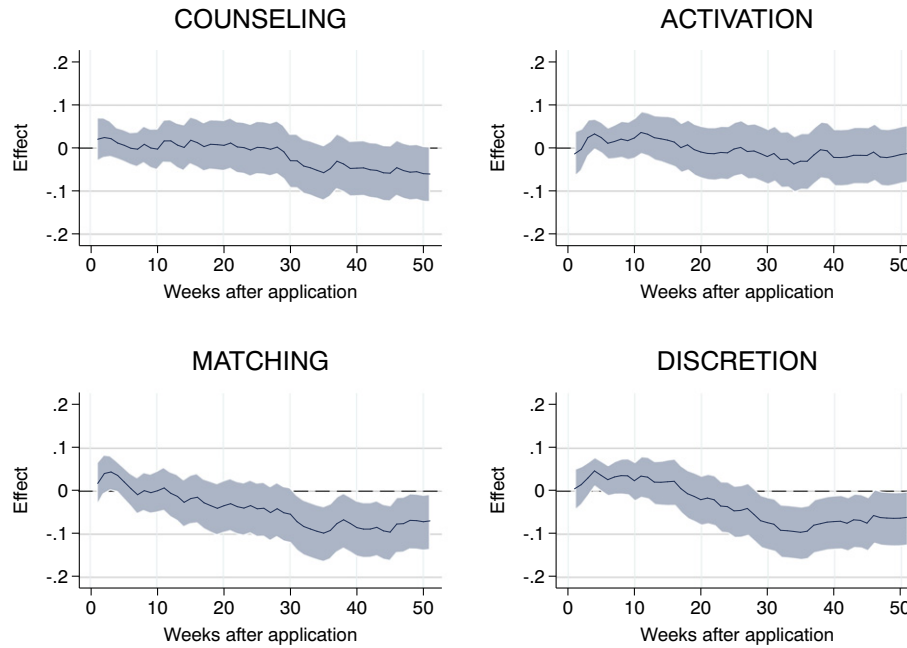
Fig. 4. Effects of policy regimes on receiving welfare benefits (0/1). Note: This figure is based on 52 separate regressions, following Eq. (1). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant.

Based on the intention-to-treat effects of the policy regimes on labor market outcomes, there are a few things that we can conclude. First, the policy regimes in which caseworkers have unrestricted access to direct matching (direct matching and full discretion) perform significantly better than no support by caseworkers with both higher exit from welfare and higher job finding. Second, if caseworkers target programs optimally in the policy regime with full discretion, this policy regime should outperform all other programs.

Based on the results in the previous section, this does not seem the case as the policy regime with direct matching performs at least as good. Third, the policy regime counseling does not lead to a higher outflow out of welfare benefits or higher earnings, compared to no support. Finally, the encouragement to participate in the job-search activation program, makes labor market outcomes significantly worse than any of the other policy regimes, including giving no guidance.

Fig. 5. Effects of policy regimes on having a job (0/1). Note: This figure is based on 52 separate regressions, following Eq. (1). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant.
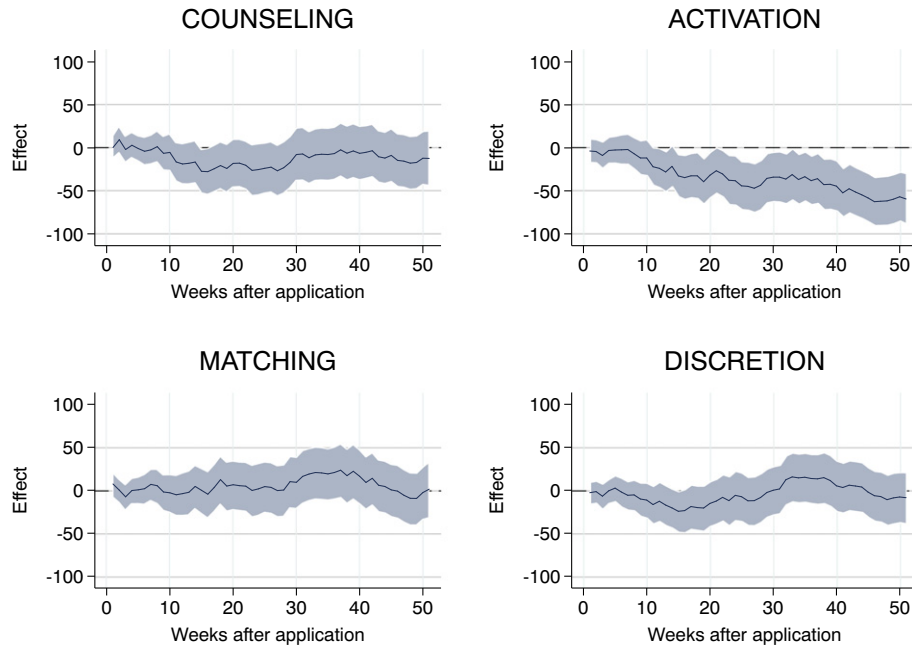
Fig. 6. Effects of policy regimes on earnings. Note: This figure is based on 52 separate regressions, following Eq. (1). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant.

5.2. Caseworker compliance and instrumental variables results

The results presented so far are intention-to-treat effects comparing the policy regimes to the policy regime without support. Compliance to the policy regimes is not perfect due to three reasons. First, caseworkers were given the option to deviate from the approach described in the policy regime. Second, in some cases welfare recipients did not satisfy the criteria for participating in a welfare-to-work program (e.g. language and computer skills for job-search activation). And third, some welfare recipients found work quickly, so before the support of the caseworker could have started. Table 3 shows which welfare-to-work programs were applied in the different policy regimes. The numbers in each column do not add up to one, as an individual can participate in multiple programs. In addition to the three welfare-to-work programs (counseling, direct matching and job-search activation) considered in the experiment,
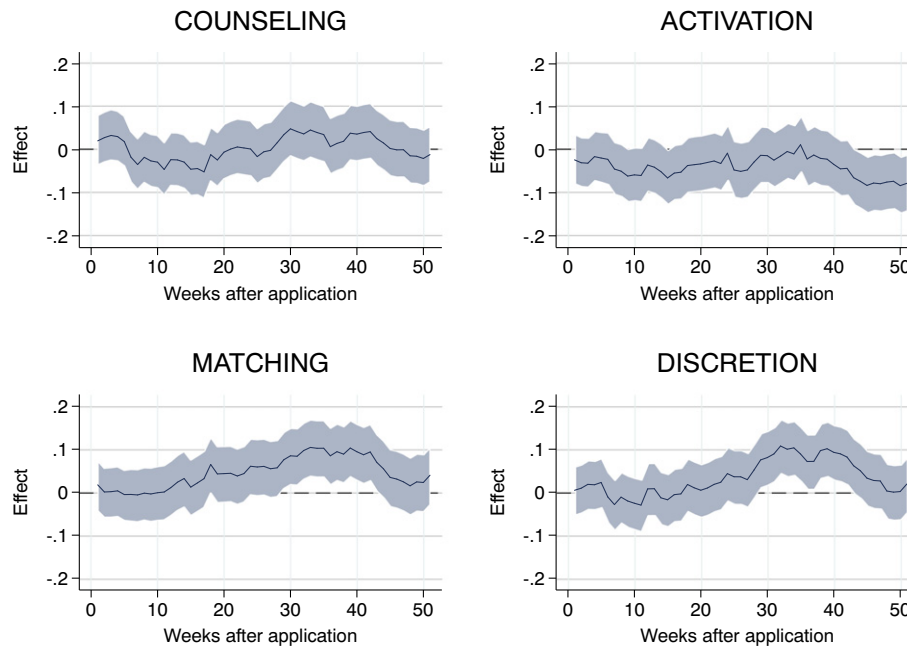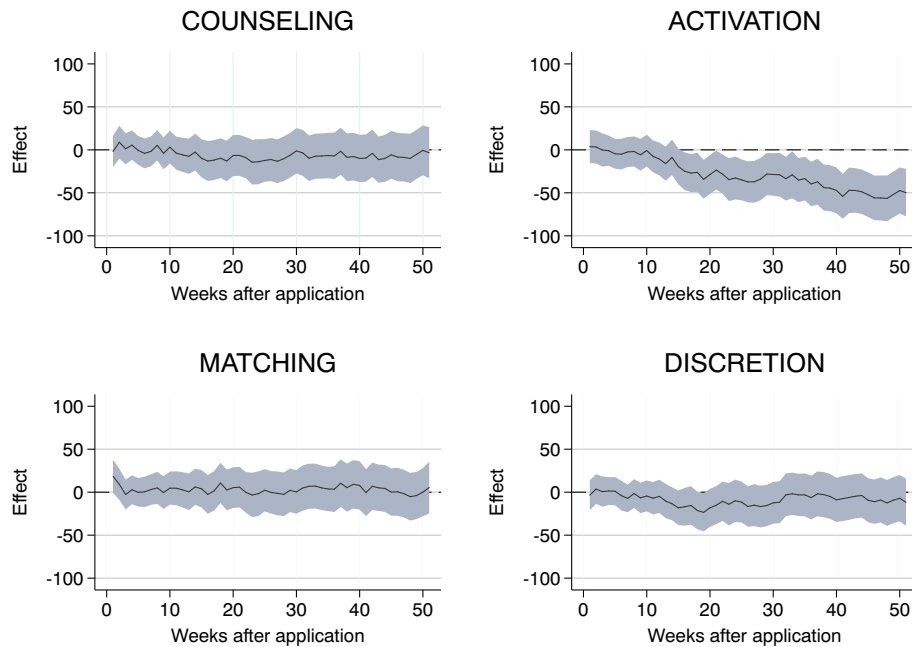

Fig. 7. Effects of policy regimes on total income. Note: This figure is based on 52 separate regressions, following Eq. (1). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant.

Table 3
Actual program participation in the different policy regimes.

| | Policy regime | | | | |
	Nothing (1)	Counseling (2)	Activation (3)	Matching (4)	Discretion (5)
No support	84%	23%	21%	29%	29%
Welfare-to-work program:					
Counseling	10%	70%	47%	49%	55%
Job-search activation	3%	6%	36%	4%	8%
Direct matching	10%	30%	27%	50%	36%
Job-hunter	2%	10%	8%	8%	10%
Other	4%	10%	19%	18%	14%
Contacts with caseworker (monthly)	0.73	1.07	1.02	1.16	0.96
N	271	353	333	328	394

we also distinguish 'job-hunter' and 'other'. The job-hunter is a program that was first introduced when our experiment was already running.[20] We could, therefore, not include it in the description of the policy regimes. The same holds for some small and very specific welfare-to-work programs that are included in the category 'other'.

The first thing to note in Table 3 is that compliance to the policy regime nothing is high. About 84% of the welfare recipients in this policy regime did not receive any support from their caseworker. This percentage is much higher than in the other policy regimes, where no support is often associated with individuals who exit welfare quickly.[21] Also the newly introduced programs (job-hunter and other) were less frequently provided to welfare recipients in the policy regime without guidance.

Looking at the other policy regimes, we see that the random assignment of the policy regimes largely increased usage of the encouraged programs compared to the regime with no support. The increase in the encouraged program use is 60, 30 and 40 percentage points for respectively counseling, job-search activation and direct matching. Interesting to note is how caseworkers allocate welfare-to-work programs in case they have full discretion, which is shown in the fifth column in Table 3. In that case about 30% of the welfare recipients do not receive any support, 55% receive counseling, less than 10% participate in the job-search activation program and direct matching is applied to 36 %. Each program is used less than in case of the policy regimes with encouragement for the program, but each program is used much more than in the policy regime without any support. The program choice in the policy regime with full discretion has quite some similarities with direct matching, maybe with the exception that the job-search activation program is used twice as frequent and direct matching somewhat less frequent. If we also take the job-hunter and other programs into account, welfare recipients that do receive support, on average, participate in two welfare-to-work programs.

The bottom row of Table 3 shows the average number of times that the caseworker contacts the welfare recipient per month under the different policy regimes. Under no guidance, the caseworkers had less contact with the applicant than under the other policy regimes. The number of meetings under no guidance is not zero because the caseworkers occasionally have to meet their clients in order to

discuss the technical aspects of their benefits payments. However, we asked the caseworkers in this policy regime not to be proactive towards the benefits recipient with respect to job coaching.

To learn more about the added value of the various welfare-to-work programs, we apply an instrumental variables approach. We regress the outcome $Y_{i\tau tw}$ of individual i observed t periods after applying for welfare benefits at local office w at calendar time τ, on actual participation in counseling ($\tilde{C}_{i\tau w}$), job-search activation ($\tilde{A}_{i\tau w}$) and direct matching ($\tilde{M}_{i\tau w}$)

$$Y_{i\tau tw} = \alpha_{\tau t} + \gamma_{wt} + \delta_t^c \tilde{C}_{i\tau w} + \delta_t^a \tilde{A}_{i\tau w} + \delta_t^m \tilde{M}_{i\tau w} + X_i \beta_t + U_{i\tau tw} \quad (2)$$

Since actual participation in a program is selective, we use the randomly assigned policy regimes as instrumental variables, which gives the first-stage regression equations,

$$\tilde{C}_{i\tau w} = \kappa_\tau^c + \lambda_w^c + \psi^{cc} C_{i\tau w} + \psi^{ca} A_{i\tau w} + \psi^{cm} M_{i\tau w} + \psi^{cd} D_{i\tau w} + X_i \phi^c + V_{i\tau w}^c$$
$$\tilde{A}_{i\tau w} = \kappa_\tau^a + \lambda_w^a + \psi^{ac} C_{i\tau w} + \psi^{aa} A_{i\tau w} + \psi^{am} M_{i\tau w} + \psi^{ad} D_{i\tau w} + X_i \phi^a + V_{i\tau w}^a$$
$$\tilde{M}_{i\tau w} = \kappa_\tau^m + \lambda_w^m + \psi^{mc} C_{i\tau w} + \psi^{ma} A_{i\tau w} + \psi^{mm} M_{i\tau w} + \psi^{md} D_{i\tau w} + X_i \phi^m + V_{i\tau w}^m$$
$$(3)$$

Since there are five policy regimes, we have four instrumental variables, while there are only three welfare-to-work programs. Above we showed that there are substantial differences in how frequent the difference programs are used in each policy regime. This is also reflected in the first-stage regressions which show significant differences in program participation between the policy regimes (see Table A4 in the Appendix). In all specifications the F-statistic for joint significance of the policy regimes on the uptake of the different programs are sufficiently high.

Most programs start between eight and 16 weeks, but some welfare recipients start a program later. This means that the classification of an individual as never taker, complier or always taker can depend on the realized period of welfare benefits receipt.[22] If the policy regimes not only affect the participation in programs but also the moment of starting the program, the instrumental variables estimates are biased away from zero and defining compliers is more complicated. Furthermore, to interpret the estimated effects from the instrumental variables approach usually a monotonicity

---

[20] A job-hunter is a person (not the caseworker) that has to acquire vacancies and find suitable welfare recipients to fill these vacancies. As such, it is somewhat similar to the instrument direct matching, the difference being that the initiative for the match comes from the job-hunter instead of from the caseworker/welfare recipient and that the job-hunter does not use trial periods or wage subsidies.

[21] Fig. A1 in the Appendix shows the likelihood to receive no guidance against the months until exiting welfare. For all policy regimes excluding no guidance, more than 60 percent of the individuals that exit welfare within two months after registration receive no guidance. This reduces to around ten percent conditional on exit after ten months. Only for the policy regime nothing the percentage of individuals without support is consistently around 80%.

[22] Consider, for example, an individual who benefits a lot from direct matching. If this individual is assigned to the policy regime with direct matching, she will enter direct matching quickly and find work quickly. If this individual is assigned to another policy regime, she may at a much later stage of the benefits period still participate in direct matching and then find work. This specific individual would be an always taker, but the potential treated outcomes depend on the policy regime.

Table 4
Effects of policy regimes and program participation on cumulative labor market outcomes 52 weeks after application.

	Weeks on welfare		Weeks worked		Earnings		Income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Policy regimes (ITT)								
Counseling	-0.95	-1.26	0.18	0.36	-633	-806	-319	-703
	(1.46)	(1.36)	(1.57)	(1.59)	(683)	(679)	(614)	(605)
Activation	-0.23	-0.76	-2.10	-2.05	-1793***	-1932***	-1521***	-1798***
	(1.50)	(1.40)	(1.58)	(1.58)	(654)	(638)	(587)	(575)
Matching	-2.34	-2.67*	2.35	2.06	262	-155	159	-315
	(1.50)	(1.41)	(1.60)	(1.63)	(693)	(676)	(608)	(597)
Discretion	-1.76	-3.14**	1.70	2.03	-240	-248	-455	-777
	(1.42)	(1.34)	(1.56)	(1.58)	(655)	(642)	(577)	(572)
Programs (LATE)								
Counseling	0.12	-0.14	-1.58	-0.93	-1535	-1377	-775	-1054
	(2.93)	(2.62)	(3.14)	(2.97)	(1307)	(1194)	(1175)	(1067)
Activation	1.86	1.05	-8.04**	-7.97**	-4726***	-4634***	-4331***	-4235***
	(3.87)	(3.55)	(4.00)	(3.84)	(1514)	(1380)	(1426)	(1308)
Matching	-6.76	-7.82*	8.81*	7.17	2392	1326	963	121
	(4.95)	(4.48)	(5.25)	(5.01)	(2147)	(1935)	(1899)	(1721)
Observations	1679	1549	1679	1549	1679	1549	1679	1549

Note: The top panel is based on Eq. (1); the lower panel follows Eq. (2). Columns (1), (3), (5) and (7) are based on the full sample, while columns (2), (4), (6) and (8) are conditional on the applicant being on welfare for at least eight weeks. Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant). *** =significant at 1% level, ** =at 5% level, * =at 10% level.

assumption is made (Angrist et al., 1996).[23] Monotonicity assumes that when the policy regime changes, program participation can either only increase or decrease. Finally, since we have four instrumental variables and only three welfare-to-work programs, the IV estimates for each welfare-to-work program describe a mix of compliers. For example, the IV estimate for activation relates to both compliers to the police regimes activation and to the policy regime caseworker discretion.[24] These limitations should be kept in mind when interpreting the estimated effects from the instrumental variables approach as local average treatment effects (LATE).

Table 4 shows the intention-to-treat effects (upper panel) and local average treatment effects (lower panel) for cumulative outcomes 52 weeks after registration.[25] The table shows the estimated effects for the full sample (odd columns) and for the subsample that collects welfare benefits for at least eight weeks (even columns). This sample selection should not affect the estimated local average treatment effects but will increase compliance (see Table A4 in the Appendix). Individuals who leave the benefits system within eight weeks usually do not participate in programs and are thus never takers.

The intention-to-treat effects confirm the earlier findings, the job-search activation regime significantly reduces earnings and income and for direct matching and full discretion exit from welfare benefits increases. The local average treatment effects show that participation in the job-search activation program increases the welfare benefits period only very modestly, but reduces the weeks worked substantially. This causes that participants in this program experience more weeks without any income, which significantly reduces total earnings and total income. The job-search activation program seems ill suited for the welfare recipients in our experiment, which have relatively favorable labor market prospects among welfare recipients. We explain the latter from the relatively long waiting period to enter the activation program and the strong focus on the temp work sector, which is more often found not to be a stepping stone for unemployed

job seekers (Autor and Houseman, 2010; Van der Klaauw and Ziegler, 2019). Participation in direct matching causes that individuals collect welfare benefits for fewer weeks and have more weeks of paid work. Although not significantly, cumulative earnings and cumulative income both increase. Counseling does not have any positive effects on job findings and earnings. This suggests that without providing vacancies the assistance of caseworkers is not very useful.

Columns (2), (4), (6) and (8) of Table 4 show the results when we restrict the sample to individuals who receive welfare benefits for at least eight weeks. As expected (because we mainly exclude never takers) the results for this restricted sample are very similar to the full sample results. Finally, Table A5 in the Appendix presents the estimates for cumulative outcomes three years after registration. These estimates are obtained using data from Statistics Netherlands and show that the estimated effects after 52 weeks are persistent and even become somewhat larger. As mentioned earlier, the data from Statistics Netherlands are noisier and more aggregated over time periods. Given the similarity of the results we will focus mainly on the first 52 weeks in the remainder of the paper.

5.3. Heterogeneity by welfare recipient, local office and caseworker

In the previous subsection we provided average effects. However, participation in a welfare-to-work program may have different effects for welfare recipients with different characteristics. Between local offices the policy to assign welfare recipients to the various programs may differ and also within local offices caseworkers may have different approaches. This causes that samples of compliers differ between caseworkers and local offices. In this subsection we focus on heterogeneity in program effects and investigate if caseworkers allocate welfare recipients to programs based on these observables when they have discretion.

To investigate heterogeneity in the program effects, we construct a vulnerability index, that ranks individuals based on the first component of a principal component analysis using a rich set of pre-experiment covariates.[26] The vulnerability index is normalized such

---

[23] We maintain the assumption that within local offices applicants are randomly assigned to policy regimes (exogeneity). This assumption has been discussed above and is satisfied by the experimental design.
[24] Note that this also means that the sign of the intention-to-treat effects is not necessarily the same as the sign of the local average treatment effect.
[25] Figs. A2–A5 in the Appendix show the instrumental variable results by week since application.

[26] The components are earnings in the two years before registration, welfare benefits in the two years before application, four dummies for different age categories, gender, partner, children, the reason for entry into welfare benefits, eight dummies for level of education, financial situation, low self-reliance and the expected number of months it will take to find a job, where the last three components are based on reports by the caseworker.

that it runs from zero to one. As can be seen from the estimation results this index is a strong predictor of labor market outcomes.

Table 5 presents the heterogeneous treatment effects. Again, the upper panel contains the intention-to-treat effects and the lower panel the local average treatment effects. The intention-to-treat effects show some evidence that the most vulnerable individuals suffer most from being assigned to the policy regime without support. There is, however, no strong evidence for heterogeneous treatment effects. We only find marginally significant estimates showing that participation in the job-search activation program does not have adverse earnings and income effects for the most vulnerable welfare recipients. This is in agreement with the original intention of the job-search activation program to assist the more disadvantaged welfare recipients in their job search.

Table 6 shows how often caseworkers in the four local offices use the different programs in the full discretion regime compared to the regime without support, by vulnerability of the applicant. These estimates can be considered as part of the first-stage regressions in the instrumental variables approach and are indicative of whether caseworkers allocate programs based on the observed characteristics of the applicant when they have discretion. The regressions show that under full discretion the more vulnerable welfare recipients get more support in the local offices in North and South/West and less

Table 5

Effects of policy regimes and program participation on cumulative labor market outcomes by vulnerability of the applicant.

	Weeks on welfare (1)	Weeks worked (2)	Earnings (3)	Income (4)
Policy regimes (ITT)				
Counseling	4.91	−3.70	−1996	−871
	(3.65)	(3.96)	(1794)	(1550)
Activation	1.58	−8.48**	−4495**	−3598**
	(3.81)	(4.05)	(1768)	(1562)
Matching	1.45	−1.87	−410	678
	(3.69)	(4.06)	(1862)	(1608)
Discretion	1.45	−2.68	−1937	−1386
	(3.59)	(3.96)	(1780)	(1575)
Counseling X vuln.	−12.19*	7.21	2419	884
	(6.96)	(7.95)	(3139)	(2752)
Activation X vuln.	−3.53	13.65*	5808*	4518
	(7.00)	(7.85)	(3069)	(2773)
Matching X vuln.	−8.05	9.02	1381	−1193
	(6.95)	(7.85)	(3233)	(2855)
Discretion X vuln.	−6.67	9.19	3556	1958
	(6.72)	(7.62)	(3044)	(2747)
Vulnerability index	18.42***	−20.03***	−8404***	−4366**
	(5.55)	(6.34)	(2534)	(2188)
Programs (LATE)				
Counseling	12.90	−8.01	−6525	−5105
	(11.12)	(11.99)	(5557)	(4857)
Activation	−7.39	−24.84*	−14048**	−14445**
	(14.39)	(15.07)	(6588)	(6195)
Matching	−8.41	5.25	7023	8780
	(16.94)	(18.56)	(8368)	(7316)
Counseling X vuln.	−24.76	11.67	9325	8220
	(20.22)	(22.37)	(9613)	(8462)
Activation X vuln.	16.97	34.25	19487*	21106**
	(24.18)	(26.08)	(10803)	(10392)
Matching X vuln.	2.19	10.14	−8255	−15309
	(32.36)	(36.09)	(15115)	(13459)
Vulnerability index	21.68***	−23.41***	−8666**	−3574
	(7.80)	(9.00)	(3556)	(3138)
Observations	1679	1679	1679	1679

Note: The top panel is based on Eq. (1) while the lower panel follows Eq. (2). The policy regimes (treatments) are interacted with the continuous vulnerability index, where a higher index corresponds to higher vulnerability. Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant. *** =significant at 1% level, ** =at 5% level, ∗ =at 10% level.

Table 6

Effects on program participation of being assigned to full discretion compared to regime with no support by vulnerability of the applicant and welfare office.

		Treatments	
	Counseling (1)	Activation (2)	Matching (3)
Office North:			
Discretion	0.16	−0.00	0.16
	(0.14)	(0.08)	(0.17)
Discretion X vuln.	0.88***	0.01	0.36
	(0.26)	(0.19)	(0.34)
Office Center/East:			
Discretion	0.67***	0.12**	0.32***
	(0.13)	(0.06)	(0.10)
Discretion X vuln.	−0.62**	−0.12	−0.24
	(0.28)	(0.12)	(0.22)
Office New West:			
Discretion	0.19	0.27**	−0.05
	(0.23)	(0.12)	(0.21)
Discretion X vuln.	−0.06	−0.60*	−0.19
	(0.48)	(0.34)	(0.47)
Office South/West:			
Discretion	0.33**	0.00	0.52***
	(0.16)	(0.08)	(0.15)
Discretion X vuln.	0.69**	0.13	−0.09
	(0.30)	(0.17)	(0.30)

Note: This table reports the first stage estimates by welfare office for the effect of the policy regime discretion on treatment uptake following Eq. (3), where all policy regimes included and are interacted with the continuous vulnerability index. A higher vulnerability index corresponds to higher vulnerability. The number of observations by welfare office are 303 (North), 534 (Center/East), 272 (New West) and 570 (South/West). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant. *** =significant at 1% level, ** =at 5% level, ∗ =at 10% level.

support in Center/East and New West. In the latter two offices, caseworkers are also more likely to apply job-search activation to *less* vulnerable applicants. Both actions are opposite to what would be recommended based on the results from Table 5.

The results indicate that the different local offices employ different policies towards supporting welfare recipients when they have full discretion. For example, North and South/West hardly apply job-search activation and New West almost never uses direct matching. These results show that in the usual regime welfare recipients are likely to get different support in different local offices. Of course, this can also (partly) reflect that local labor markets and populations of welfare recipients differ between local offices.

Differences in how programs are applied are stronger between caseworkers than between local offices. Fig. 8 shows for each caseworker the fraction of welfare recipients that participate in a particular program in the policy regime of full discretion. There is substantial dispersion between caseworkers in how often they assign welfare recipients to the different welfare-to-work programs. For example, some caseworkers almost never use direct matching while others use it for the majority of their welfare recipients. Overall rates of use for the job-search activation are low and many caseworkers never assign welfare recipients to this training, but a few caseworkers very often assign welfare recipients to this program.

We use the information from Fig. 8 to define specialization by caseworkers. In particular, we distinguish between caseworkers who use a welfare-to-work program more or less frequent than the mean of all caseworkers. Next, we interact the indicator for frequent use with the assigned policy regime. This provides insight in whether caseworkers that specialize in a particular welfare-to-work program obtain better outcomes with this program. It is not ex-ante clear if specialization is positive or not. Caseworkers that specialize may become better in using a program, but they may also assign too many
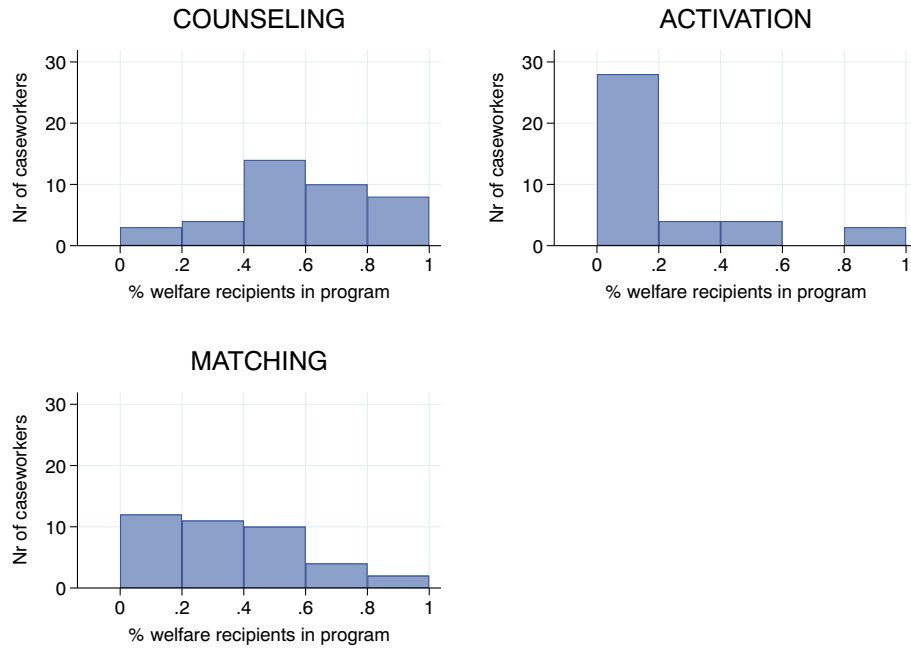
Fig. 8. Fraction of welfare recipients that participate in each welfare-to-work program by caseworker in full discretion regime.

welfare recipients to this program and do not obtain enough information about the other available programs. It is thus an empirical question how the effects of programs differ between caseworkers that specialize and do not specialize.

Table 7 presents the estimated local average treatment effects of participating in the different welfare-to-work programs on the different cumulative labor market outcomes (in the year after application). The rows labeled with 'low use' provide the effects for caseworkers that have a low use of the program under full discretion, while the rows with 'high use' present the estimated effects for caseworkers with an above average use of the program. The estimation results do not provide evidence that the effect of participating in a program depends on how much experience the caseworker has with the program. The estimated effects of the low-use and high-use caseworkers are similar and never significantly different from each other.

Our experimental design would have the potential to estimate many marginal treatment effects, because caseworkers behave differently in case of full discretion and respond differently to the encouraged policy regimes.[27] Therefore, each caseworker generates estimated effects for a different population of compliers. However, for a caseworker the number of welfare recipients in each policy regime is not very large and the results above show that estimated effects do not vary much by program use (and thus compliance) of the caseworker. This suggests that the marginal treatment effects do not vary much. In addition, our experiment lacks the statistical power to identify small variations in the marginal treatment effects.

Instead of estimating marginal treatment effects we test if the effects of participating in a program differ between welfare recipients who are likely to participate in programs and unlikely to participate in programs. The first group would include individuals on the margin of being a never taker or a complier, while the second group would include individuals on the margin of being an always taker or complier. We construct an index describing expected program use during the period of welfare benefits receipt and split the sample

into welfare recipients with above and below median expected program participation. Next, we estimate the effects of participating in the different programs for both groups. The estimates are shown in Table 8. The results provide some weak indication that individuals who are likely to get many treatments suffer less from participating in the job-search activation program and benefit (in terms of having work instead of benefits) from direct matching. This coincides with the earlier findings using the vulnerability index. The vulnerability index correlates with expect program participation because both describe individuals who are likely to collect welfare benefits for a longer period.

5.4. Caseworker learning

Above we showed that effects of participation differ between welfare-to-work programs, but the effectiveness of a welfare-to-work program does not depend on the caseworker. Still, caseworkers often have a very different approach when supporting welfare recipients. In particular, in the usual policy regime they have substantial discretion and they use this to focus on different welfare-to-work programs. Our experiment forced caseworkers to change their usual approach and to also consider other welfare-to-work programs when supporting welfare recipients. Furthermore, the experiment provides insight in the effectiveness of the different programs, which was not systematically collected before. In this subsection we study if caseworkers respond to this by changing their attitudes for supporting welfare recipients.

After the experiment and also after we had presented results to the management and caseworkers at the local offices, we conducted a survey among caseworkers. In the survey we asked caseworkers to answer for each of the welfare-to-work programs 'Compared to before the experiment, have you changed your belief about to the usefulness of the program'. Caseworkers answered these questions on a scale from 0 to 100 where we indicated less useful (<50) no change (50) and more useful (>50). The results, which are summarized in Table 9, do not concur with our empirical findings. In particular, 43% of the caseworkers believe that direct matching is less useful, while we find it to be the most effective instrument. Similarly, 47% of the caseworkers report that counseling is more useful

---

[27] This would make the additional assumption that the caseworker does not have a direct effect on the labor market outcomes of welfare recipients.

Table 7
Effects of participating in the welfare-to-work programs on cumulative labor market outcomes by caseworker specialization (instrumental variables estimation).

	Weeks on welfare (1)	Weeks worked (2)	Total earnings (3)	Total income (4)
Counseling:				
Low use caseworker	−0.56	−3.84	−2327	−837
	(5.18)	(5.51)	(2150)	(1843)
High use caseworker	0.42	−2.55	−1923	−1054
	(3.09)	(3.36)	(1367)	(1198)
Activation:				
Low use caseworker	0.81	−7.54	−3009	−1776
	(6.29)	(6.64)	(2420)	(2224)
High use caseworker	2.01	−6.69	−4134***	−3558**
	(4.02)	(4.27)	(1597)	(1483)
Matching:				
Low use caseworker	−11.87	17.69*	5709	2631
	(8.46)	(9.24)	(3646)	(3004)
High use caseworker	−8.38	13.61**	3981*	2178
	(5.15)	(5.65)	(2266)	(1926)

Note: The two rows under the name of each policy regime represent each one regression including the interaction of the relevant policy regime with the low or high use indicator. Included controls are the low or high use indicator, indicators for the other policy regimes, calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant. *** =significant at 1% level ** =at 5% level, ∗ =at 10% level.

Table 8
Effects of participating in the welfare-to-work programs on cumulative labor market outcomes by likelihood to receive many treatments.

	Weeks on welfare (1)	Weeks worked (2)	Total earnings (3)	Total income (4)
Counseling:				
Low exp. nr treatments	−1.41	−2.32	−2383	−1237
	(4.63)	(4.89)	(2084)	(1913)
High exp. nr treatments	2.22	−0.97	−1311	−771
	(3.64)	(3.98)	(1658)	(1490)
Activation:				
Low exp. nr treatments	1.71	−13.58	−6031*	−6718**
	(8.20)	(8.54)	(3321)	(3184)
High exp. nr treatments	1.25	−5.62	−3758***	−3109**
	(3.85)	(3.91)	(1463)	(1392)
Matching:				
Low exp. nr treatments	−4.47	4.02	3669	3819
	(7.93)	(8.38)	(3715)	(3279)
High exp. nr treatments	−10.45**	10.71*	2305	−214
	(5.48)	(5.93)	(2359)	(2100)

Note: The two rows under the name of each policy regime represent each one IV regression including the interaction of the relevant policy regime with the indicator for low or high expected number of treatments. Included controls are the low or high expected number of treatments indicator, indicators for the other policy regimes, calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant. *** =significant at 1% level ** =at 5% level, ∗ =at 10% level.

than they thought before the experiment, even though we find that counseling does not outperform providing no support. Finally, 29% answer that the job-search activation program is more useful, a program that we find to be not only non-effective, but also harmful to having work and total income of the welfare recipient.

Next, we asked caseworkers in the survey if they actually changed their usual approach of supporting welfare recipients in their job search.[28] In particular, we asked them about their use of each welfare-to-work program in relation to before the experiment. The answers are again on a scale from 0 to 100, which we categorized as less use (<50), no change (50) in use and more use (>50). The results are presented in Table 10. The tendency is to intensify the use of welfare-to-work programs, and in particular counseling. The results also show that more caseworkers want to use job-search activation more often than less frequently, and direct matching is mentioned most often as program that will be used less frequently.

When asked whether the caseworkers found the experiment useful only five out of 50 answered negatively. The responses to the survey indicate that a majority of the caseworkers provide support

differently after the experiment. However, the caseworkers do not use empirical evidence to update their beliefs about the welfare-to-work programs. Their answers about the effectiveness of programs and their intended use are not aligned with our empirical evidence on the effectiveness of the different programs. A likely explanation is that due to the experiment caseworkers learn about other characteristics of the welfare-to-work programs which they value more than effectiveness. For example, some caseworkers indicated that they do not like direct matching because it provides fully subsidized labor to employers, who have only limited commitment.

Table 9
Perceived effectiveness of different programs compared to beliefs prior to the experiment.

	Less useful (1)	No change (2)	More useful (3)
Counseling	19%	33%	47%
Job-search activation	34%	37%	x29%
Direct matching	43%	11%	46%
N	36		

Note: The information in this table is based on the post-experiment survey among the caseworkers that participated in the experiment. The response rate to the survey is 81%.

---

[28] Due to data limitations we do not observe the welfare-to-work programs that caseworkers applied after the experimental period.

Table 10
Use of different programs compared to before the experiment.

	Use less (1)	No change (2)	Use more (3)
Counseling	11%	33%	56%
Job-search activation	23%	46%	31%
Direct matching	31%	34%	34%
N	36		

Note: The information in this table is based on the post-experiment survey among the caseworkers that participated in the experiment. The response rate to the survey is 81%.

6. Conclusion

In this paper we study the role of the caseworker in the assignment and take-up of active labor market programs. We conduct a field experiment where five different policy regimes are randomly assigned to caseworkers and welfare recipients are randomly assigned to caseworkers. We find that direct matching to vacancies is effective in increasing exit from benefits by helping welfare recipients to find paid work. Participation in the job-search activation program makes labor market outcomes significantly worse than giving no guidance to the welfare recipient. We do not find strong evidence in favor of heterogeneous treatment effects. The more disadvantaged welfare recipients benefit somewhat more from participation in the welfare-to-work programs, but program participation effects do not vary between caseworker. Still, we find substantial differences in the use of programs between local offices and even more between caseworkers. Our post-experimental survey shows that some caseworkers continue focusing on job-search activation, even after being informed that it has adverse consequences on the labor market prospects of benefits recipients.

Our findings imply that even when there is credible knowledge about the effectiveness of certain active labor market programs, the roll-out of such a program is not obvious. Caseworkers have substantial discretion, and do not easily change behavior. This finding confirms earlier results of Behncke et al. (2009) and Schmieder and Trenkle (2016). While in their case caseworkers were only provided with small incentives to change behavior, we show that this is still the case if caseworkers are provided with extensive information about the effectiveness of different programs and experience the use of these programs. As a response, benefits agencies can formulate policy regimes to reduce discretion, for example by shutting down some active labor market programs completely or enforcing mandatory participation for some groups of benefits recipients. But even in such regimes the heterogeneity of benefits recipients requires some discretion for caseworkers. Overall, our results show that learning which active labor market programs are effective is not enough to increase the use of these programs.

A. Appendix

Table A1
Benefit levels (net, in €per month).

	Full	Housing costs Shared	None
Single without children	935.80	802.12	668.43
Single with children	1203.19	1069.50	935.81
Couple without children	1336.87	1203.19	1069.50
Couple with children	1336.87	1203.19	1069.50

Note: Benefit levels in period July 1 to December 31 in 2012, including holiday allowance. Benefit levels outside this time frame differ only marginally. Shared housing costs apply if the costs are shared with an individual that is not the partner or the child.

Table A2
Description of the seven fictional clients in the post-experiment survey.

Name	Description
Person 1	A married man of 50 years old. He has limited command of Dutch, and only finished primary school.
Person 2	A woman of 38 years old. She is divorced and has two children aged four and seven. Her ex-partner refuses to pay the alimony, so she hardly manages to make ends meet. She has little work experience.
Person 3	A single man of 33 years old. The past few years he only had temporary jobs. He only followed preparatory vocational education, but never finished his vocational degree.
Person 4	A highly educated woman of 51 years old. She has a lot of work experience and has never relied on welfare benefits before. She has made a lot of job applications, but is continuously rejected.
Person 5	A man of 45 years old. He applies for welfare benefits because his shop went bankrupt. He wants to have a regular job and has already deregistered as self-employed at the chamber of commerce (Kamer van Koophandel).
Person 6	A single woman of 32 years old. She has a lot of debts, rent, phone bills, health insurance, mail order companies, etc. At the first meeting she arrives late.
Person 7	A single man of 29 years old. He graduated last year and traveled afterwards. Now he is looking for his dream job.

Table A3
Log of visits to the local welfare offices.

Name	Description
March 2012	Information meetings at welfare offices.
March 2012	Visits and collection of forms.
April 2012	Information meetings at welfare offices.
April 2012	Visits and collection of forms.
April - May 2012	Sit in with caseworker meetings, visits and collection of forms at all welfare offices.
May 2012	Visit to statistical office, first data delivery.
June 2012	Handout of new default options at all welfare offices.
June 2012	Progress report to all team managers.
July 2012	Visit and collection of forms at all welfare offices.
September 2012	Handout of new default options and collection of forms all welfare offices.
November–December 2012	Presentation of progress experiment at all welfare offices.
December 2012	Handout of new default options and lists of missing forms at all welfare offices.
January 2013	Collection of forms and completed lists of missing forms at all welfare offices.
February 2013	Collection of forms at all welfare offices.
March 2013	Collection of forms at all welfare offices.
April 1, 2013	End of experiment.
April 2013	Second data delivery.
April–May 2013	Presentation of (preliminary) results at all welfare offices.
October 2013	Survey sent out to all participating caseworkers.
November 2013	Sample of the experiment matched with data from the social security administration.
February 2014	Presentation of final report (incl. results of the survey). Policy makers, management teams and caseworkers invited.

Table A4
First-stage effects of default options on treatment uptake.

| | Treatments | | | | | |
| | Counseling | | Activation | | Matching | |
	(1)	(2)	(3)	(4)	(5)	(6)
Policy regimes						
Counseling	0.61***	0.66***	0.01	0.01	0.17***	0.19***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Activation	0.39***	0.42***	0.32***	0.33***	0.13***	0.15***
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Matching	0.40***	0.43***	-0.01	-0.01	0.37***	0.40***
	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.04)
Discretion	0.46***	0.48***	0.05***	0.05**	0.25***	0.27***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
F-statistic	114.9	123.8	37.4	35.7	34.0	34.9
Observations	1679	1549	1679	1549	1679	1549

Note: This table reports the first stage estimates for the effect of the policy regimes on treatment uptake following Eq. (3). Columns (1), (3), (5) and (7) are based on the full sample, while columns (2), (4), (6) and (8) are conditional on the applicant being on welfare for at least eight weeks. The policy regime with no support is the base category. Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). The *F*-statistics describe joint significance of the instrumental variables in each first-stage regression. Standard errors are clustered at the level of the applicant. *** =significant at 1% level, ** =at 5% level, ∗ =at 10% level.

Table A5
Effects of policy regimes on cumulative labor market outcomes three years after application (data from Statistics Netherlands).

	Weeks on welfare (1)	Weeks worked (2)	Total earnings (3)	Total income (4)
Policy regimes (ITT)				
Counseling	−9.03**	−0.91	−210	−1230
	(4.41)	(4.20)	(1642)	(1259)
Activation	−2.07	0.76	−1857	−1597
	(4.40)	(4.26)	(1611)	(1239)
Matching	−10.99**	4.80	2645	647
	(4.37)	(4.18)	(1670)	(1314)
Discretion	−8.88**	4.48	525	−1345
	(4.28)	(4.08)	(1544)	(1157)
Programs (LATE)				
Counseling	−9.98	−6.75	−3175	−3701
	(8.99)	(8.41)	(3359)	(2590)
Activation	13.72	1.91	−6470	−2947
	(11.58)	(11.13)	(4158)	(3256)
Matching	−18.91	22.97*	10102*	4710
	(14.79)	(13.95)	(5650)	(4488)
Observations	1679	1679	1679	1679

Note:The top panel is based on Eq. (1); the lower panel follows Eq. (2). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant. *** =significant at 1% level, ** =at 5% level, ∗ =at 10% level.
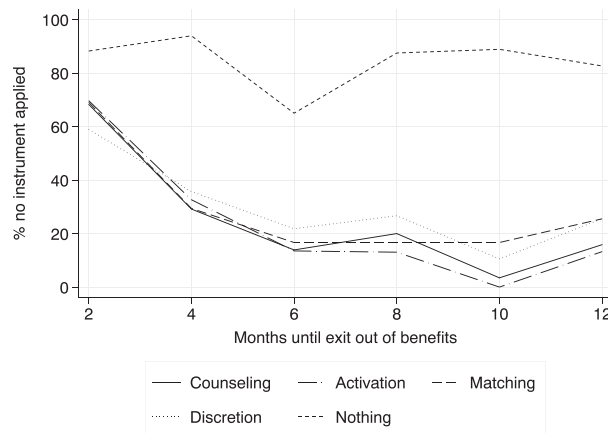

Fig. A1. No instrument applied to client, by duration of the benefits spell.

COUNSELING
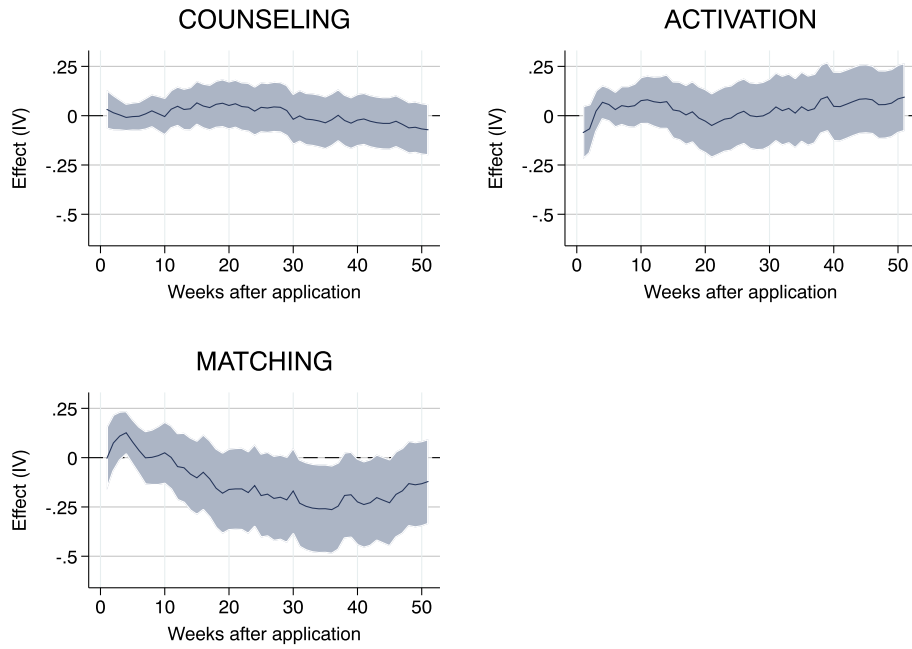
ACTIVATION

MATCHING

Fig. A2. Effects (IV) of treatments on receiving welfare benefits (0/1). Note: This figure is based on 52 separate regressions, following Eq. (2). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant.
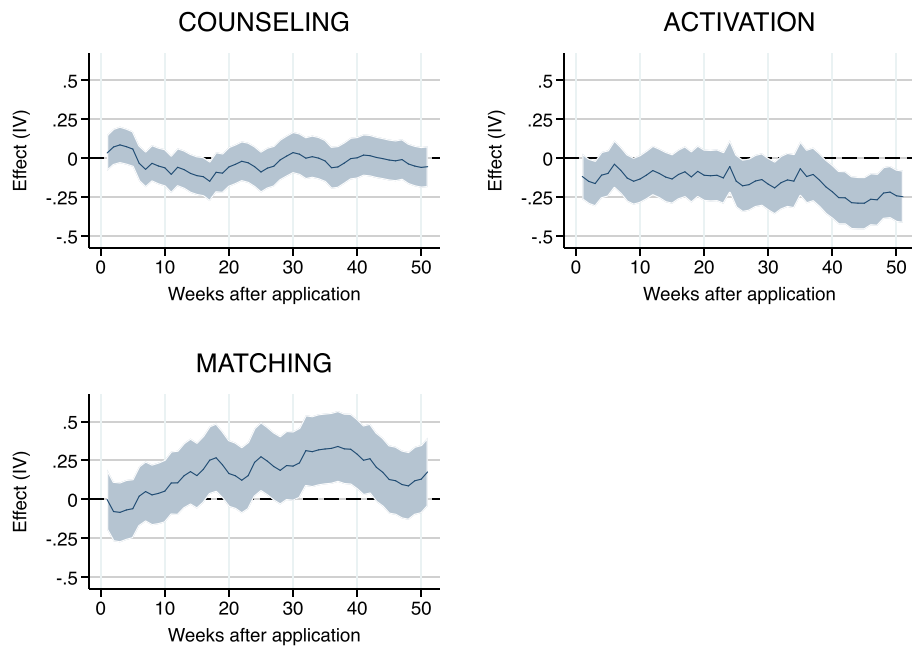
COUNSELING

ACTIVATION

MATCHING

Fig. A3. Effects (IV) of treatments on having a job (0/1). Note: This figure is based on 52 separate regressions, following Eq. (2). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant.
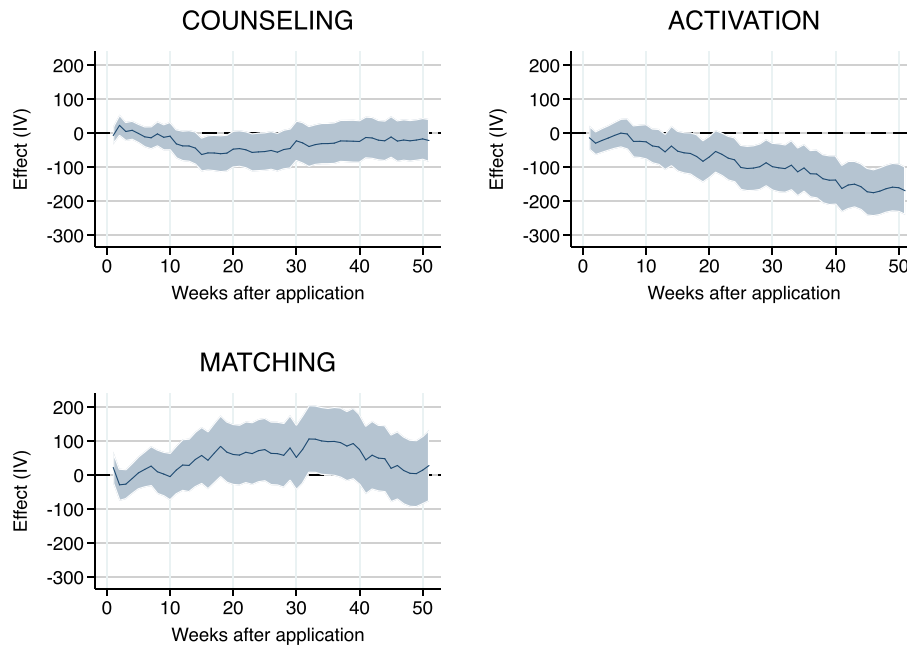
COUNSELING

ACTIVATION

MATCHING

Fig. A4. Effects (IV) of treatments on earnings. Note: This figure is based on 52 separate regressions, following equation (2). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant.
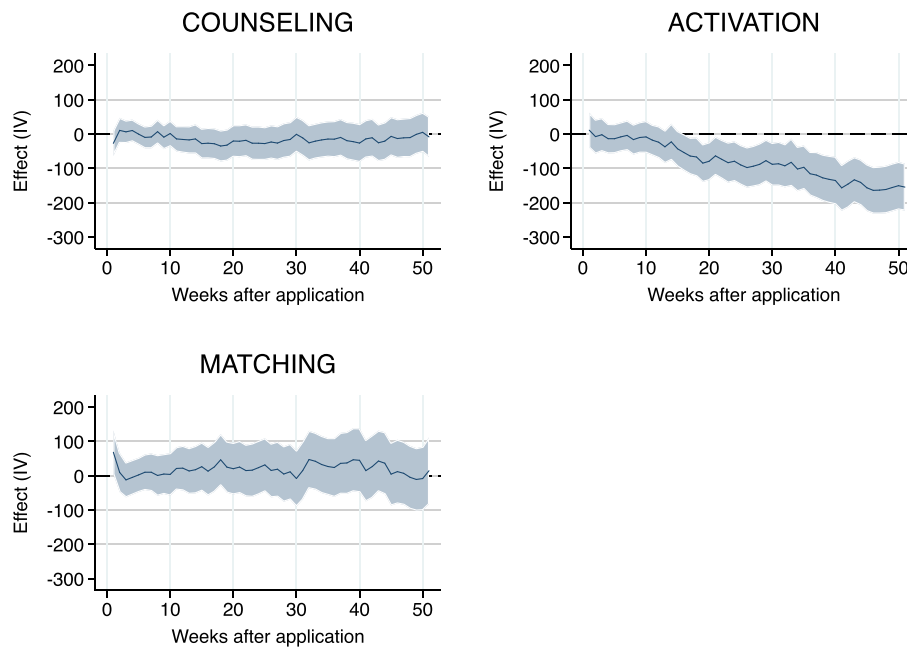
COUNSELING

ACTIVATION

MATCHING

Fig. A5. Effects (IV) of treatments on total income. Note: This figure is based on 52 separate regressions, following Eq. (2). Included controls are calendar time fixed effects, local office fixed effects and applicant characteristics (age at registration, gender, household composition, cumulative income in the 24 months before registration and dummies for five education categories). Standard errors are clustered at the level of the applicant.

References

Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. J. Am. Stat. Assoc. 91 (434), 444–455.

Autor, D.H., Houseman, S.N., 2010. Do temporary-help jobs improve labor market outcomes for low-skilled workers? Evidence from 'work first'. Am. Econ. J.: Appl. Econ. 2 (3), 96–128.

Behncke, S., Frölich, M., Lechner, M., 2009. Targeting labour market programmes - results from a randomized experiment. Swiss J. Econ. Stat. (SJES) 145 (III), 221–268.

Bell, S., Orr, L., 2002, April. Screening (and creaming?) applicants to job training programs: The AFDC homemaker - home health aide demonstrations. Labour Econ. 9 (2), 279–301.

Bolhaar, J., Ketel, N., van der Klaauw, B., 2019. Job-search periods for welfare applicants: evidence from a randomized experiment. Am. Econ. J.: Appl. Econ. 11 (1), 92–125.

Card, D., Kluve, J., Weber, A., 2010. Active labour market policy evaluations: a meta–analysis. Econ. J. 120 (548), F452–F477.

Card, D., Kluve, J., Weber, A., 2018. What works? A meta analysis of recent active labor market program evaluations. J. Eur. Econ. Assoc. 16 (3), 894–931.

Dahl, G.B., Kostol, A.R., Mogstad, M., 2014. Family welfare cultures. Q. J. Econ. 129 (4), 1711–1752.

Eriksson, M., 1997. To Choose or not to Choose: Choice and Choice Set Models (Ph.D. thesis), Umeå University.

Graversen, B.K., Van Ours, J., 2008. How to help unemployed find jobs quickly: experimental evidence from a mandatory activation program. J. Public Econ. 92 (10&11), 2020–2035.

Heckman, J.J., Lalonde, R.J., Smith, J.A., 1999. The economics and econometrics of active labor market programs. Handbook of Labor Economics, 3. Elsevier, pp. 1865–2097.

Huber, M., Lechner, M., Mellace, G., 2017. Why do tougher caseworkers increase employment? The role of program assignment as a causal mechanism. Rev. Econ. Stat. 99 (1), 180–183.

Lechner, M., Smith, J., 2007. What is the value added by caseworkers? Labour Econ. 14 (2), 135–151.

Maestas, N., Mullen, K.J., Strand, A., 2013. Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. Am. Econ. Rev. 103 (5), 1797–1829.

Rosholm, M., 2014. Do case workers help the unemployed? IZA World Labor 72,

Rothstein, J., Von Wachter, T., 2017. Social experiments in the labor market. In: Banerjee, A.V., Duflo, E. (Eds.), Handbook of Economic Field Experiments. 2. pp. 555–637.

Schiprowski, A., 2020. The role of caseworkers in unemployment insurance: evidence from unplanned absences. J. Labor Econ. Forthcoming,

Schmieder, J.F., Trenkle, S., 2016. Disincentive effects of unemployment benefits and the role of caseworkers. J. Public Econ. Forthcoming. (April).

Van den Berg, G.J., Van der Klaauw, B., 2006. Counseling and monitoring of unemployed workers: theory and evidence from a controlled social experiment. Int. Econ. Rev. 47 (3), 895–936.

Van der Klaauw, B., Ziegler, L., 2019. A field experiment on labor market speeddates for unemployed workers. Discussion Paper, CEPR.

Vikström, J., 2017, December. Dynamic treatment assignment and evaluation of active labor market policies. Labour Econ. 49, 42–54.