// NO.19-048 | 11/2019

DISCUSSION PAPER

// BORIS IVANOV, FRIEDHELM PFEIFFER, AND LAURA POHLAN

Do Job Creation Schemes Improve the Social Integration and Well-Being of the Long-Term Unemployed?

Do job creation schemes improve the social integration and well-being of the long-term unemployed?*

Boris Ivanov[1)], Friedhelm Pfeiffer[2)] and Laura Pohlan[3)]

[1)] ZEW Mannheim[†]
[2)] ZEW Mannheim, University of Mannheim
[3)] ZEW Mannheim, IAB Nuremberg

October 2019

Abstract

In this paper we analyze the effects of a German job creation scheme (JCS) on the social integration and well-being of long-term unemployed individuals. Using linked survey and administrative data for participants and a group of matched non-participants, we find significant positive effects of being employed within this program. They are larger for individuals with health impairments and above-average duration of welfare dependence. The program effects decline over time, which cannot be explained by decreasing levels of well-being and social integration of the participants. Instead, this decrease is driven by a rising share of controls who find a job and catch up to similar outcome levels as program participants. Overall, our results suggest that JCSs can be an efficient labor market policy instrument to improve the quality of life of the long-term unemployed.

Keywords: unemployment, active labor market policy, job creation schemes, well-being, social integration, matching
JEL-Code: I31, J64

---

*The authors are part of the ongoing research project "Evaluation of the Federal Program ‘Social Integration within the Labor Market’" financed by the Federal Ministry of Labour and Social Affairs. The linked survey and administrative dataset ‘SILM Evaluation Dataset’ used in this paper, was developed as part of this research project. We sincerely thank our cooperation partners at the Institute for Work, Skills and Training in Duisburg, Zoom - Society for Prospective Developments in Göttingen and SOKO Institute for Social Research and Communication in Bielefeld as well as Bernd Fitzenberger, Christina Gathmann, Anthony Lepinteur, Johanna Poetzsch and participants of seminars at the University of Heidelberg, ZEW Mannheim, IAB Nuremberg and the annual conferences of ESPE, EALE and VfS in 2019 for helpful discussions. Access to administrative data provided by the IAB Nuremberg are also gratefully acknowledged. The views expressed in this study are those of the authors and do not necessarily reflect the views of the Federal Ministry of Labour and Social Affairs. All opinions and errors are our own.

†Address of correspondence: ZEW – Leibniz Centre for European Economic Research, research area of Labour Markets and Human Resources, L 7,1, D-68161 Mannheim, E-mail: boris.ivanov@zew.de, friedhelm.pfeiffer@zew.de and laura.pohlan@zew.de

1 Introduction

Exclusion from the labor market is a serious risk factor for social exclusion and reduced well-being in many Western societies (see e.g. Clark and Oswald, 1994; Paul and Moser, 2009 and Pohlan, 2019). Beyond the economic strain it puts on individuals, unemployment may cause substantial psychological and social costs. Compared to working individuals, the unemployed often suffer from a higher risk of depression, suicide, alcohol abuse and stigmatization that stresses self-esteem and personal relationships (see e.g. Frey and Stutzer, 2002). The consequences of unemployment may even persist across generations (see e.g. Clark and Lepinteur, 2019). This raises the question of how labor market policy can help to improve the quality of life of the long-term unemployed.[1] There are two basic policy options: compensation or re-integration into work.

First, unemployment insurance is meant to mitigate material hardship. However, even a generous scheme may not suffice to compensate for the negative consequences of job loss, because employment in itself has essential psychosocial functions such as social purpose, status and identity (Jahoda, 1981). Alternatively, the barriers to early retirement or disability pension schemes could be lowered, as many long-term unemployed are of advanced age or face health problems. On the one hand, retirement can have a greater identity value than unemployment (Hetschko et al., 2014). On the other hand, incentivizing work-capable individuals to leave the active labor force may come at great fiscal costs (see e.g. Autor and Duggan, 2006; Autor et al., 2016).

Second, policy could foster re-integration into the labor market. There are several options for such active policy measures, among them incentivizing job search, continuous training and job creation. Incentivizing job search and continuous training may be effective for relatively healthy and skilled unemployed individuals, but may fail for the long-term unemployed. One reason is that employers often interpret long unemployment durations as a significant negative signal (see e.g. Kroft et al., 2013; Bhuller et al., 2017). If this is the case, such re-integration policies would be ineffective on average. As a third alternative to passive welfare receipt, the government could provide subsidized job opportunities directly to individuals with employment impediments.

In this paper we use a recent German job creation scheme (JCS) to analyze whether subsidized employment improves the social integration and well-being of the long-term unemployed. In contrast to previous JCSs, the federal pilot project 'Social Integration within the Labor Market' (SILM, *Soziale Teilhabe am Arbeitsmarkt*) explicitly aimed

[1] We use the terms 'long-term unemployed' and 'welfare recipients' interchangeably, because in our application all long-term unemployed individuals are also welfare recipients.

at social integration rather than increasing the re-employment prospects.[2] So far, only a few studies have investigated whether subsidized employment can directly improve the quality of life of participants and the evidence is mixed. Crost (2016) shows that between 1992 and 2004 German JCSs improved life satisfaction of participants. In a similar vein, Knabe et al. (2017) find that participation in the more recent German job creation program 'One-Euro-Jobs' is associated with higher life satisfaction based on cross-sectional data. Using fixed effects regressions, Wulfgramm (2011) provides only weak evidence that One-Euro-Jobs can partially counteract the negative effects of unemployment on life satisfaction. Gundert and Hohendanner (2015) find that this program does not generally improve social belonging. Huber et al. (2011) estimate slightly negative effects of welfare-to-work programs on mental health in Germany, while Korpi (1997) finds positive effects for Swedish youths in the 1980s.

This study contributes to the program evaluation literature in four ways. First, we focus on long-term unemployed individuals with employment impediments. Although the target group of SILM amounts to more than 600 thousand individuals in Germany, this group is underrepresented in large scale surveys like the SOEP. Therefore, we use a novel and unusually rich dataset which links social security data of the German Federal Employment Agency with a panel survey of program participants and control individuals (see Brussig et al., 2019). Second, this paper investigates how participation in a JCS affects social belonging, social status, life satisfaction and mental health using a single dataset and methodology, where previous studies have focused on only one of these outcomes. Third, we investigate how the effects change over the course of the program and the mechanism driving this result. Fourth, to the best of our knowledge, we are the first to analyze the benefits and costs of an active labor market program using life satisfaction as an indicator of individual utility.

We apply nearest-neighbor propensity score matching to estimate the average treatment effect of SILM in three different survey waves, which roughly correspond to a duration of 7, 18 and 29 months since individual program entry. Our findings suggest that participation significantly increases well-being and social integration, but to varying degrees for the four outcome measures. While life satisfaction strongly increases, social status improves only moderately. The average program effects substantially decrease over the course of the program. This cannot be explained by decreasing outcome levels of participants, but by control individuals who catch up over time: the share of employed controls increases and they reach similar levels of well-being and social integration as program participants.

[2]Recent meta-analyses find that subsidized employment schemes, on average, do not increase the post-program employment probability (see e.g. Kluve, 2010 and Card et al., 2018).

These findings have direct policy implications: the effectiveness of the program could be increased by directing it to those with severe employment impediments, individuals with above-average duration of welfare dependence or with health impairments. Heterogeneity analyses reveal that these groups benefit more from the program. Furthermore, fixed effects estimates show that caseworker interventions and joint activities with other participants are associated with a lower drop-out probability and accompanying training courses are related to better outcomes.

The cost benefit analysis suggests that the monetary equivalent of the individual utility gains from participating in the program are at least as high as the net costs per participant for the government. This is not only because the program effect on life satisfaction is relatively high, but also because most of the net earnings from working in the program are deducted from participants' welfare claims.

The rest of the paper has the following structure. Section 2 gives an overview of the program design and its implementation. Section 3 describes the data sources. Section 4 explains how we link administrative and survey data and construct the final estimation sample. Section 5 describes the identification strategy and potential threats. Section 6 presents the estimates of the program effect, how it evolves over time, some robustness checks, its heterogeneity and the potential underlying mechanisms. Section 7 provides the cost benefit analysis. Section 8 concludes.

2 The federal program 'Social Integration within the Labor Market'

The federal job creation program 'Social Integration within the Labor Market' was launched as a pilot project to combat long-term unemployment and its consequences.[3] It offered subsidized employment for up to 20,000 participants between the fourth quarter of 2015 and the end of 2018.

In order to participate in the program, job centers had to apply for funding from the Federal Ministry of Labour and Social Affairs. Out of the 408 German job centers 265 filed an application. The ministry decided on these applications in two rounds based on criteria reflecting the local labor market situation and the job centers' implementation plans. In 2015, funding for 10,000 places was granted to 105 job centers. In 2017, 90 additional job centers were given access to the program. In this second round, 10,000

[3] In May 2015, a second JCS for hard-to-place long-term unemployed welfare recipients was launched, the so-called 'ESF-federal program' (see Boockmann et al., 2017). Although the target group of this program was similar to the one of SILM, its main emphasis was on improving re-employment prospects rather than social integration.

additional program places were allocated to both new and previously participating job centers. Participating job centers were concentrated in areas with weak economic conditions and a high share of welfare recipients. Within job centers, case workers matched individuals with employers and individual participation was voluntary.[4]

The target group of SILM are individuals who are at least 35 years of age, have been welfare claimants for at least four years and either have health impairments, minor children or both.[5] Table 1 provides selected descriptives of participant characteristics and different aspects of the program design that are derived from both administrative data and a participant panel survey.[6] Participants were on average 49 years old and had been dependent on welfare for 7.4 years. For about half of them the administrative records list some kind of health impairment as an employment impediment.[7] About one fourth of the participants lived in households with minor children.

Like previous JCSs in Germany, SILM offered subsidized employment as an ultima ratio for long-term unemployed individuals with low employability. However, this program was different in spirit: it primarily aimed at promoting social integration and well-being, whereas improving re-employment prospects was a secondary goal. This shift in focus is reflected in the specific design of the program. Employment opportunities offered by prior JCSs, such as One-Euro-Jobs, were mostly short-term and part-time. SILM, in contrast, provided up to 36 months of subsidized employment, resulting in an average planned duration of about two years (see Table 1). Participants received a regular work contract including pension claims, holiday entitlements etc. and job centers subsidized wage costs in the amount of the national minimum wage (8.50 EUR per hour in 2015 and 2016, 8.84 EUR from 2017 onwards) and the employer share of social security contributions.

The subsidy covered up to 30 working hours a week. To meet the needs of both participants and employers, SILM also allowed for different part-time agreements (15, 20, 25 hours a week) as well as a gradual increase in the number of hours worked. On average, participants worked about 28 hours per week and about 71% of all participants worked 30 hours or more. Given a 30 hour contract and the national minimum wage, the

[4]Potential selection issues due to this design are discussed in Section 5.

[5]The program aimed at welfare recipients under Book II of the Social Code (*Sozialgesetzbuch II*, SGB II), i.e. individuals who are considered capable of working and part of the labor force. Exceptions in the admission rules applied to individuals without a vocational degree and to former participants of the public employment program 'Citizen Work'. A rough estimate of the number of potential participants amounts to about 670,000 persons in November 2015. This figure was kindly provided by the Institute for Employment Research (IAB). According to this approximation, the program covered about 3% of the target population.

[6]See Sections 3 and 4 for an introduction to the data sources and the construction of the sample.

[7]It should be noted that the indicator for health impairments is based on the subjective judgment of the case workers at the job centers.

Table 1: Program design and participant assessment

	Mean
Eligibility criteria	
Age	49.31
Years of welfare receipt	7.42
Health impairment	0.52
Children in household	0.26
Program characteristics	
Planned program duration [months]	24.65
Welfare receipt	0.73
Program drop-out	0.07
Employment-accompanying activities:[§]	
Personal counseling	0.19
Training/qualification measure	0.17
Support by case worker/coach	0.18
Activities with other participants	0.08
Health promotion	0.08
Any employment-accompanying activity	0.42
Job characteristics	
Average working hours per week	27.88
Full-time employment [$\geq$30h/week]	0.71
Tasks:[§]	
Social work	0.31
Gardening/crafts/janitor	0.29
Administration/archive/library	0.11
Kitchen/food distribution	0.08
Cleaning/housekeeping	0.07
Sales/social department stores	0.12
Others	0.18
Participant assessments	
Job satisfaction [0-10]	8.27
Work is meaningful	0.92
Good relations with colleagues	0.82

Notes: The table shows the means of selected program characteristics over all three waves of the panel survey of participants (see Section 3 and 4). Eligibility criteria are based on administrative data. For details see Table A.1 and A.2 in the Appendix. The number of observations ranges from 4,917 to 7,947 due to missing values. [§] Multiple answers possible.
Source: SILM Evaluation Dataset, Brussig et al. (2019).

maximum wage subsidy amounted to 1,320 Euros per month. According to the German social security laws, the first 100 Euros of monthly earnings are exempt, but 80 to 90% of any income above that threshold is deducted from the participants' welfare claims.[8] As a result, participants earned an average additional income of about 3,350 Euros per year from working in the program. This figure also represents the additional wage costs for the government as compared to a state of pure welfare dependence. Despite working in SILM, Table 1 shows that 71% of participants still received additional welfare payments.

The tasks carried out in the program were required to be of public interest, competition neutral and additional in nature, i.e. they should not crowd out existing jobs or tasks. Most of the program places were assigned to public employers or charity organizations. Thus, the majority of participants performed tasks in social work (31%) or gardening, crafts or janitorial work (29%). Other tasks included administrative, kitchen, cleaning and sales activities.

SILM also provided different accompanying activities that were meant to stabilize the employment relationship or to foster social integration beyond employment itself. Table 1 presents the fraction of participants taking part in different types of such activities. They included individual counseling on strengths and weaknesses or personal goals (19%), support at the work place (17%) as well as training and qualification measures like computer courses or acquiring a forklift license etc. (17%), but also recreational activities with other participants (8%) or health promotions (8%). Between survey waves, 42% of all participants took part in at least one such accompanying activity and about 70% had done so until the last survey wave (about 2 to 3 years after program entry).

7% of participants dropped out of the program earlier than initially planned in each wave of the survey. The total share of drop-outs in the last survey wave amounts to 18%.[9] Table B.1 in the Appendix shows some descriptives for drop-outs. The most prominent reason for leaving the program early were health issues and mental or physical overload, followed by a mismatch between the expected and actual tasks or conflicts at the workplace. In wave 1 (within the first year after program entry), 75% of drop-outs returned to unemployment. Over the course of the program, more drop-outs transitioned into employment. In wave 3, about 24% of drop-outs were full-time employed and 20% work in part-time or marginal employment. The levels of well-being and social integra-

[8]The exact share of the deduction depends on the level of gross income, see Article 11b(2) SGB II.

[9]This figure understates the true share of program drop-outs of 25% (Brussig et al., 2019). One reason could be that the willingness to participate in the survey depends on the individuals experience with the program. Consistently, we find that being a drop-out in the previous wave is a predictor of non-response in the follow-up wave. We account for that when using the inverse response probabilities in a weighted regression as a robustness check (see Section 6).

tion of drop-outs are lower than for participants and similar to those of non-participants.

Table 1 shows that participants rated the program very positively. Their job satisfaction amounted to 8.27 on a 0 to 10 scale. This may be due to the quality of social interactions at work and a perception that the tasks they carry out were appreciated by others. For example, 82% enjoyed good relations with their colleagues and superiors and 92% perceived their work as meaningful.

3 Data sources

Although welfare recipients are a group of considerable size and of particular interest to policy makers, they are typically underrepresented in datasets that contain measures of well-being or social integration. For that reason, high quality administrative data were augmented with a telephone survey of the program's target group. While participants and potential controls can be identified in the administrative data, the survey data provides us with information on well-being and social integration and usually unobserved details of the individual level program design.

The administrative data source, the *Integrated Employment Biographies (IEB)* of the German Federal Employment Agency, is based on employer notifications to the social security system. They are available for all persons with at least one entry in their social security records after 1975 for West Germany and 1992 in East Germany. Periods of self-employment, civil service or military service are not included in the dataset. The IEB contain detailed information on individual employment histories, including spells of dependent employment, registered unemployment, job-search or benefit receipt periods on a daily basis. This enables us to construct detailed measures of the number of spells and durations of different employment states throughout the employment biography for each individual. Moreover, accurate information on the length and frequency of participation in active labor market policy (ALMP) programs is available. The data further includes information on (daily) wage records and firm characteristics such as industry code, median wage or firm size. We also include information on the local labor market situation provided by the Statistical Office of the Federal Employment Agency.

The administrative data is supplemented by a longitudinal *telephone survey* of program participants and control individuals, that provides different measures of well-being and social integration.[10] Our outcome variables include two measures of subjective well-being: life satisfaction and mental health. The variable life satisfaction is based on a question which is standard in large-scale surveys like the SOEP or the European Value

[10]We call the linked survey and administrative dataset 'SILM Evaluation Dataset'.

Study. Individuals are asked to assess current satisfaction with their life in general on a 0 to 10 scale, with 0 meaning completely dissatisfied" and 10 meaning "completely satisfied". To measure mental health, respondents were asked to asses how often they had been struggling with problems like e.g. fear, dejection, irritability or insomnia in the past four weeks. Here, the scale ranges from 1 "all the time" to 5 "never".

Social integration is measured in two dimensions: the subjective perceptions of social belonging and social status. Social belonging refers to the feeling of being part of society, which is measured on a 1 to 10 scale, where 1 means "I feel excluded" and 10 means "I feel included". To measure social status, individuals were asked to rank themselves on a 1 to 10 scale, where 1 means belonging to "the bottom" of society and 10 to "the very top".

To make the results comparable to other studies, our outcome measures were chosen to match the variables in the PASS dataset.[11] In addition to the outcome variables, the survey data contain detailed information on the particular implementation of the program at the individual level (e.g. tasks performed, working hours, participation in activities accompanying employment and subjective evaluations).

4 Sample construction

The construction of our final estimation sample involved five basic steps that are closely related to the empirical identification strategy. In steps 1 to 3 participants and control individuals were identified and matched in the administrative data. In step 4 the resulting treated and control group enter the telephone survey and in step 5 the final estimation sample is obtained. Table A.3 in the Appendix gives an overview of this process and how the observation numbers evolve over these five steps briefly described in this section.

Step 1: Identification of participants. As a first step, all early program participants (November 2015 to June 2016) and a random draw of 3,500 late entrants (after the expansion of the program in January 2017) were identified in the administrative data. This results in a sample of 12,412 individuals.

Step 2: Identification of potential control individuals. For each participant, 20 non-participants were pre-selected to match them along the eligibility criteria of SILM (age, duration of benefit receipt, incidence of health impairments, minor children) and gender.

[11]The PASS covers 21,000 individuals from low-income households in Germany, see Trappmann et al. (2010). The same measure of social belonging is used by Gundert and Hohendanner (2014; 2015). Pohlan (2019) analyzes the effects of unemployment on the same four outcomes.

Step 3: Matching. Then nearest-neighbor propensity score matching with replacement was applied to narrow this sample down to the four most comparable controls for each participant. The propensity of program participation was estimated by probit models using a detailed set of pre-treatment individual, job, firm and regional characteristics and information from the entire employment and ALMP history (see Table A.1 in the Appendix for a detailed description of all explanatory variables).[12] These individuals form the gross sample for the first wave of the telephone survey.

Step 4: Telephone survey. For *wave 1* of the survey, all program participants were contacted as soon as possible to obtain a measurement of the outcomes at an early phase of program participation.[13] This leads to a sample of 3,821 participants, which corresponds to a response rate of 30.8%.[14] The mean program duration in wave 1 amounts to 7 months. If a participant responded to wave 1, the survey institute tried to reach at least one of the four nearest neighbors in the control group.[15] This leads to a sample of 3,451 controls in wave 1. On average, they are surveyed about 4 months after their matched participant.

These successfully surveyed matches entered the survey pool of *wave 2*, which is obtained by a similar procedure: once a program participant from wave 1 responded to wave 2, his or her matched control was called again. The response rate of participants in wave 2 was 71% and the average elapsed time since program entry is 18 months. In the control group, the response rate was slightly lower (63%).

All program participants who answered the survey in wave 2 are contacted again in *wave 3*.[16] The response rate was 51% and the time since program entry amounts on average to 29 months. In the control group the response rate amounts to 41%.

Step 5: Final estimation sample of matched pairs. In a last step, the estimation sample was restricted to matched pairs, for which both sides have been successfully surveyed and all outcome and control variables are non-missing. Matches with control individuals who enter the ESF-federal program are excluded because they are subject to an alternative treatment. This last step results in a sample of 2,529 matched pairs in

[12]Since nearest-neighbor matching was carried out with replacement, some control individuals were assigned to several participants. The matched control group thus consists of 31,131 individuals.

[13]It should be noted that participants only appear in the administrative data once they enter the program, such that it was not possible to conduct a pre-treatment survey of the outcomes. We will discuss how this limitation affects our empirical strategy in the following section.

[14]This response rate is in line with other surveys in comparable populations like the PASS (Berg et al., 2016).

[15]If none of them responded, the scope was widened to the 10 nearest neighbors. This does not affect the quality of the matches since it turns out that the estimated propensity scores of participants and their first 10 matches are very close.

[16]For participants who entered the program after the expansion in January 2017 only two waves were conducted.

wave 1, 1,171 in wave 2 and 443 in wave 3. We use this unbalanced panel of matched pairs to estimate the program effects for waves 1 to 3, which correspond to an average program duration of about half a year, one and a half years and two and a half years.[17] The waves are referred to as t_1, t_2 and t_3 in the following analyses.[18]

5 Empirical strategy

5.1 Empirical model

We estimate the average treatment effect on the treated (ATT) of SILM over the course of the program. We do so by comparing the group means of program participants and matched non-participants while controlling for differences in observables using data from three survey waves, that represent different program durations. This procedure can be summarized with the following estimation equation

$$y_{i,t} = \sum_{s=1}^{3} (\beta_s Treat_i \times t_{i,s}) + X_{i,t}\gamma + \varepsilon_{i,t}, \tag{1}$$

where $Treat_i$ is an indicator of program participation for individual $i = 1, ..., N$ and $t_{i,s}$ is an indicator for wave $s = 1,2,3$. y_{it} represents the outcome variables for individual i in wave $t = 1,2,3$. $X_{i,t}$ is a set of pre-treatment control variables (for a description see Tables A.1 and A.2 in the Appendix), $\varepsilon_{i,t}$ is the idiosyncratic error term. To make the estimated effect sizes comparable across outcomes, the outcome variables are standardized to a mean of zero and unit standard deviation. The parameters of interest are β_s, which identify the ATT in waves 1 to 3 in terms of standard deviations of the outcome variables.

When estimating the effect of SILM, it has to be taken into account that selection might arise at three different levels: (i) job center selection into the program, (ii) individual selection into the program and (iii) individual selection into the survey. In the following subsection, we will discuss the implications of each selection problem and how we address them.

[17] We also estimate the program effects based on the balanced panel of matches (see Section 6.3).

[18] Figure B.1 in the Appendix plots the distribution of the outcome variables and their means and standard deviations in the final sample. Table B.4 provides mean values of pre-treatment characteristics and outcomes by wave and treatment status.

5.2 Selection issues

(i) Job center selection into the program. Not all job centers applied for funding under SILM. Hence, participating and non-participating job centers might differ with respect to characteristics that may also affect well-being and social integration. Participating job centers are, in general, more concentrated in regions with higher unemployment rates and bleak employment prospects for welfare recipients. Worse economic conditions are likely to be related to well-being and social integration, but the direction of a potential bias is not clear.[19] To account for such differences at the job center level, we make use of a classification of job centers developed by Dauth et al. (2013) when estimating the propensity score of program participation. This classification reflects disparities in the local economic and social structure that affect the strategies of local job centers. In a robustness check, we compare the estimates of matches with control individuals from both participating and non-participating job centers.

(ii) Individual selection into the program. In addition, there is individual level selection within job centers: program entry is voluntary and expert interviews at job centers suggest that case workers tried to ensure a good fit to a particular job or employer when considering potential participants (Brussig et al., 2019). As a consequence, unobserved characteristics like soft-skills and competencies or motivation and ability to work might be determinants of program entry. At the same time, they are likely correlated with pre-treatment outcome levels. Again, the direction of potential bias is not clear a priori.[20]

Individuals selection into the program is addressed in the sample construction process by applying nearest-neighbor propensity score matching to choose the most comparable control individual for each participant. We use the detailed employment and ALMP records in the administrative data to construct different proxy measures for unobservable time invariant confounders.[21] Especially information about previous experience in ALMP programs might be informative about selection into the current program as well. Prior participation in One-Euro-Jobs turns out to be a strong predictor of partic-

[19] Since a higher general unemployment rate is associated with lower levels of individual well-being, participating job centers might be negatively selected with respect to overall well-being (Di Tella et al., 2001). This would bias our estimated ATT upwards. On the other hand, individual well-being is positively associated with reference group unemployment (Clark, 2003). Being one of many welfare recipients in a participating job center might thus have less of a negative reputation effect, such that welfare recipients in participating job centers might be positively selected and our estimated ATT would be biased downwards.

[20] On the one hand, individuals with greater well-being might be more motivated to enter the program causing an upward bias in our estimated ATT. On the other hand, program participants might ascribe greater value to employment and have suffered more from unemployment prior to entering the program, which would result in a downward bias.

[21] Caliendo et al. (2017) show that including explicit measures for personality traits, expectations, social networks etc. does not change the estimated impact of ALMP on employment if labor market histories are controlled for.

ipation in SILM. Both are likely related to time-invariant unobservables like willingness to work, attitudes towards the case worker and job centers' support opportunities. We also include the share of unfinished ALMP programs as a proxy measure for motivation, endurance or capability.

The first wave of the survey provides some characteristics that can be regarded as reflecting the pre-treatment living conditions and emotional state of respondents. In particular, they were asked whether they had ever felt the need for support in different life domains such as caring for minor children, psychological problems and addiction or indebtedness. These are used as control variables in equation 1 to account for differences that were still unobserved when the matching was carried out. As a further robustness check, we include survey data reflecting an individual's willingness to work as additional control variables.

To check the quality of the matching procedure, we assess whether the common support assumption is fulfilled and whether observable characteristics are balanced between participants and controls in our final estimation sample of wave 1. Figure B.2 in the Appendix shows that propensity scores of treated and controls overlap almost perfectly. Table 2 compares selected sample means between both groups. Participants and controls are similar with respect to most of these characteristics. The only exception is a significant difference in the years of welfare dependence. Considering the average duration of around 7 years of welfare receipt, a difference of 2.5 months seems to be of negligible importance for the analysis. Table 2 also shows the aforementioned indicators from the survey measuring individuals' need for support. In each group, respondents most prominently mentioned psychosocial or addiction problems and indebtedness. Although there are significant differences between both groups, they seem to face similar problems in many areas of life. We take this as evidence that the matching approach to some extent accounts for characteristics related to individual life circumstances. Nevertheless, the regressions include pre-treatment control variables both from administrative and survey data to account for any imbalances in observables that remain after matching.

(iii) Selection into the survey. Selection into the telephone survey may also affect the estimates. Suppose, for example, individuals with lower pre-treatment outcomes benefit more from the program but are less likely to answer the telephone survey. This would lead to an underestimation of the ATT. Similarly, panel mortality might be non-random.

To address selection into the first survey, we first check whether the composition of the participant sample changes over the steps of the sample construction process. Table B.2 in the Appendix shows the mean values of pre-treatment characteristics and outcome variables of participants in the administrative data, in the survey sample of

Table 2: Comparison of participants and matched controls in final estimation sample for wave 1

	(1)	(2)	(3)	
	Means			
	Participants	Non-participants	Difference	
Sociodemographics				
Female	0.47	0.45	0.02	
Age	48.92	48.77	0.15	
Health impairment	0.50	0.48	0.02	
Children in household	0.27	0.28	-0.01	
German	0.93	0.93	-0.01	
Married	0.25	0.25	0.00	
No professional degree	0.17	0.17	0.00	
Vocational degree	0.78	0.79	-0.01	
Academic degree	0.05	0.04	0.01	
Region with weak employment prospects§	0.79	0.77	0.01	
Employment history				
Years of welfare receipt	7.43	7.21	0.22	**
Cum. years of ssc. employment	6.53	6.68	-0.15	
Cum. years of marg. employment	1.27	1.35	-0.07	
Cum. no of ALMP measures	6.65	6.81	-0.16	
Prior participation in "One-Euro-Job"	0.77	0.75	0.01	
Share of unfinished ALMP measures†	0.12	0.12	0.01	
Need for support				
Care for minor children	0.09	0.07	0.02	***
Psychological problems or addiction	0.15	0.16	-0.01	
Indebtedness	0.20	0.17	0.02	**
Other life domains	0.21	0.28	-0.07	***
At least one field	0.44	0.48	-0.03	**
Observations	2,529	2,529		

Notes: The table shows means of selected pre-treatment characteristics of participants (column (1)) and matched non-participants (column (2)) in the final estimation sample (see Section 4). Column (3) shows the differences in means and their significance levels from two-sample t-tests. Sociodemographics and the employment history are based on administrative data. The questions regarding the need for support are measured in survey wave 1. For details see Table A.1 and A.2 in the Appendix. § Defined as job center regions of SGB II-Type 3 according to the classification of the Federal Employment Agency (Dauth et al., 2013). † An ALMP spell is defined as unfinished if it ends before the planned end date without a transition into employment. ***/**/* marks significance at the 1/5/10% level.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

wave 1 and in the final sample of matches in wave 1. Although observation numbers decline substantially, there is no evidence of a change in composition. Turning to the control individuals, they have very similar levels of well-being and social integration as individuals in the PASS dataset who would also fulfill the eligibility criteria of SILM (compare columns (2) and (4) in Table B.3 in the Appendix). This provides evidence that non-response did not result in a systematic selection of the control individuals and that they are representative in terms of the outcome variables.

To address selection into the follow-up surveys, Table B.4 in the Appendix shows observable characteristics of participants and their controls in the final sample of matches for wave 1 to 3. The descriptives suggest that panel mortality does not lead to a different composition of surveyed individuals over time. As a robustness check, we estimate the ATTs using the balanced panel sample only. Moreover, we reweigh the estimation equation by the individual probability of program participants to respond to the first and follow-up waves of the survey.

6 Results

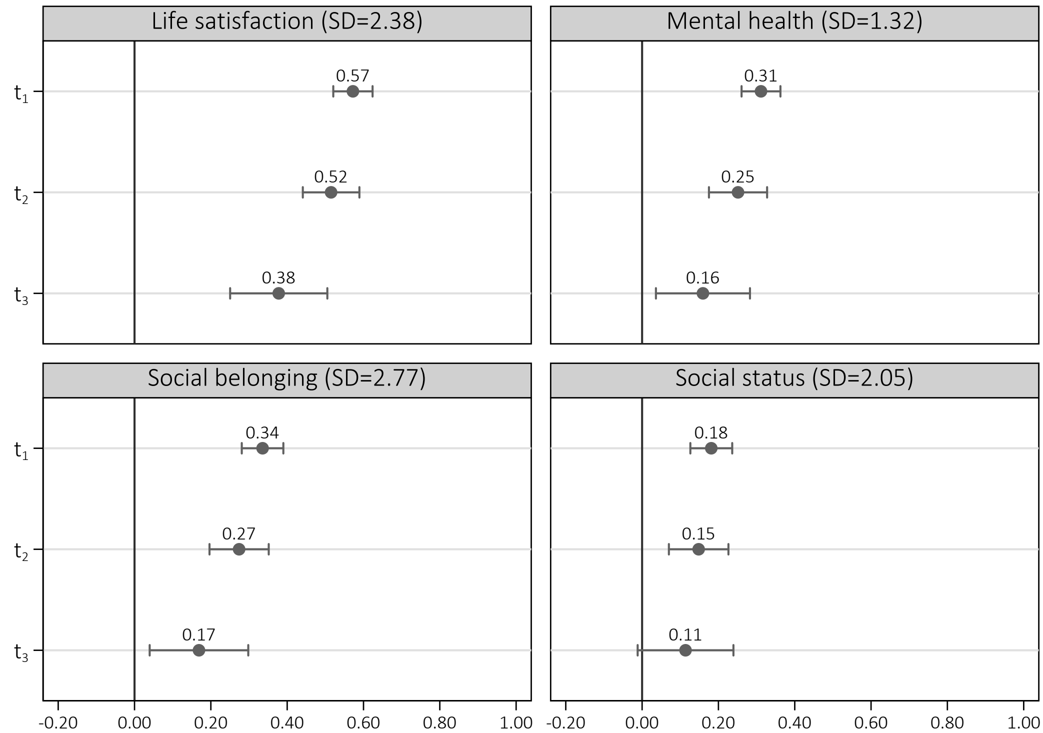
6.1 Estimated average program effects

Figure 1 depicts the estimated ATTs on our four outcome variables for the three survey waves. In wave 1, the largest program effect is on life satisfaction which increases by 0.57 standard deviations (SDs). The impact on mental health amounts to 0.31 SDs. The program effects are 0.34 SDs on social belonging and 0.18 SDs on social status.

To put these estimates into perspective, the mean levels of the outcome variables of participants are compared to different subsamples of the PASS (see Table B.3 in the Appendix). SILM increases participants' life satisfaction to the level of employed individuals in PASS, whereas social belonging and especially social status are still lower. Life satisfaction is likely to summarize economic, psychological, social and other benefits of program participation, which might explain the large size of the program effect. Being employed in the program increases social status, but the average level is lower than the average of regularly employed individuals. This might be due to the fact that working contracts in SILM are only temporary and participants earn the minimum wage.

Our estimated ATTs are similar to Pohlan (2019), who studies the causal short-term effects of job loss on the same well-being and social integration measures based on PASS. She finds effect sizes of -0.55 SDs on life satisfaction, -0.31 on mental health, -0.34 on social belonging and -0.25 on social status. While our results confirm the finding that JCSs partially counteract the negative effects of unemployment on life

Figure 1: Estimated ATTs on standardized outcomes over the course of the program



Notes: SD: standard deviation. The figure shows the estimated ATTs at different program durations (mean duration of 7 months in t_1, 18 months in t_2, 29 months in t_3) from pooled OLS regressions (see equation 1) based on the final estimation sample (see Section 4). Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). The total number of observations amounts to 8,286 (5,058 in t_1, 2,342 in t_2, 886 in t_3). Standard errors are clustered at the individual level. Whiskers represent 95% confidence intervals.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

satisfaction (see e.g. Knabe et al., 2017) and social belonging (see e.g. Gundert and Hohendanner, 2015), our estimated effect sizes for wave 1 are larger. Moreover, our positive estimate for mental health contrasts Huber et al. (2011), who find that program participation may even moderately increase the prevalence of mental symptoms and sleeplessness among welfare recipients.

The larger effects may be due to the specific target group and design of SILM. Participants are characterized by considerable barriers to the regular labor market and very long unemployment durations. As they face a particularly high risk of social exclusion, they may also have a greater scope for improvement than the average unemployed person. In contrast to previous programs, SILM offered a working contract with social security contributions and a relatively long duration. SILM explicitly aimed at improving social integration and well-being. Qualitative evidence suggests that case workers made considerable efforts to find jobs that fit both the individual participants' and employers' needs (Brussig et al., 2019). In line with this, surveyed participants consistently reported that case workers are supportive and that their jobs fit their experience and allow them to develop their skills, while only very few perceive work as too (un-)demanding or stressful.[22] Hence, the larger positive effect of participating in SILM might be explained by it's explicit focus on improving the social integration of participants and specific aspects of its design. This will be examined more closely in Subsection 6.4.

6.2 Declining program effects

Figure 1 shows that the large effects of wave 1 substantially decrease over the course of the program. In wave 3, the ATTs on life satisfaction declines to 0.38 SDs and the effect on mental health to 0.16 SD. For social belonging, the ATT goes down to 0.17 SDs. The drop is smaller for social status, but the estimated effect of 0.11 SDs in wave 3 is not significant at the 5% level. These reductions in the program effects range from -33% for life satisfaction to -48% for mental health and social belonging.

How can the decline in program effects be explained? Do outcome levels of participants go down or do control individuals catch up over time? Both mechanisms would have different policy conclusions: if participants eventually return to the baseline level of pure welfare dependence, there would be no rationale to fund long-term JCSs. If, on the other hand, control individuals catch up in absence of the program, this might have implications for the design of the eligibility criteria. Declining outcome levels in the participant group could have several reasons. First, there might be habituation. After an initial improvement, participants might get used to working in the program and

[22] These insights are not reported and are available on request.

gradually return to their baseline levels of well-being and perceived social integration. Second, with the end of the program approaching, stress and fear of returning to unemployment could dampen the program effect. Third, an increasing share of drop-outs with lower average levels of well-being and social integration could reduce the average outcomes in the participant group.

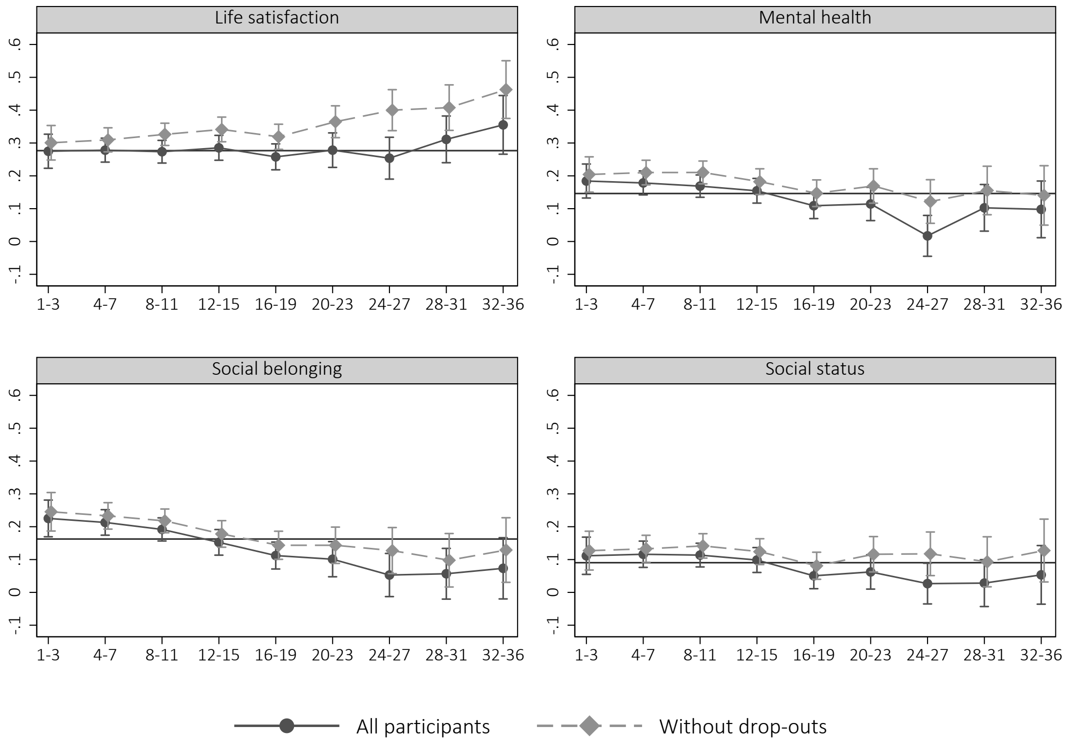
To further investigate these possible explanations, we make use of the fact that not all participants start the program at the same date and not all of them are surveyed at the exact same time after entry. As a consequence, there is some variation in the elapsed time since program entry within each survey wave. We divide the elapsed time into 3-month-intervals and predict the mean outcome levels within each of them.[23] The results are displayed in Figure 2. There is a slightly negative trend for mental health, social belonging and social status – although the latter two do not significantly deviate from the mean over all time intervals (horizontal line). When drop-outs are excluded, these trends vanish. Active participants do not seem to face habituation or fear of the program ending, as their outcome levels remain rather stable over the course of the program, life satisfaction even increases. For all participants, there is only weak evidence that well-being and social integration of participants revert over the course of the program. An increasing share of drop-outs slightly drags down average well-being and social integration in the participant group. However, the impact of drop-outs is not large enough to explain the substantial decrease in the ATTs as shown in Figure 1.

Given stable outcome levels in the treatment group, the fall in the ATTs may be explained by the control group catching up. Over the course of the program, the share of control individuals who find employment increases: in wave 1, 26% of the control individuals in our estimation sample are employed – about 17% have a full-time job. Until wave 3, the share of employed controls further increases to 36%. These individuals reach similar levels of well-being and social integration as program participants. To show this, we predict the average outcomes of employed and unemployed controls separately for each wave and compare them to the levels of participants.[24] Figure 3 plots the resulting conditional means. It shows that unemployed controls face considerably lower levels of well-being and social integration than participants. In contrast, employed control individuals reach similar levels as program participants. Over the course of the

[23]We regress the outcome variables on a set of indicators of the elapsed time intervals and the same set of pre-treatment control variables as in equation 1 (see Tables A.1 and A.2 in the Appendix). These estimates are based on the pooled survey data of all three waves. Therefore, the intervals ranging from month 1 to 15 roughly correspond to observations of wave 1, the intervals of 16 to 23 months to wave 2 and the intervals of 24 to 36 months to wave 3.

[24]These predictions are derived from a similar model as in equation 1 with additional interaction terms of an indicator for employment and wave indicators.

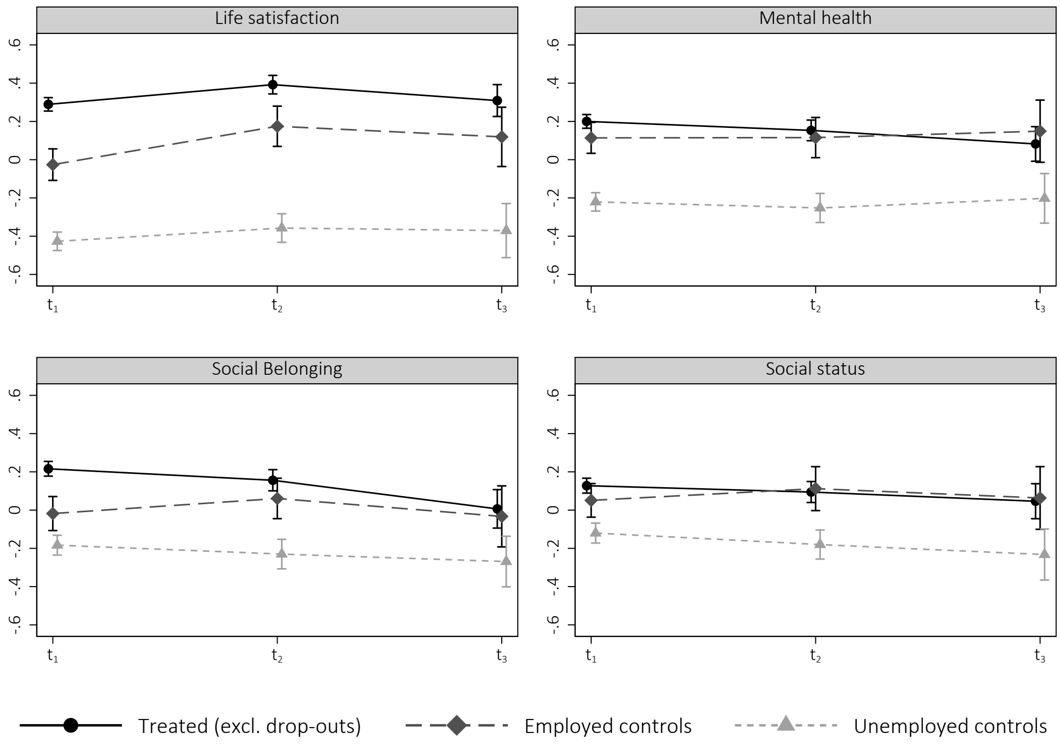
Figure 2: Outcomes of participants by time spent in program, with and without drop-outs



Notes: The figure shows the conditional means of the outcome variables of participants by time since program entry within 3-month-intervals based on the final estimation sample (see Section 4). The horizontal line marks the average over all intervals. Conditional means are predicted by OLS regressions of the standardized outcomes variables on indicators of the time intervals. Control variables contain an indicator for the entry cohort (before or after 2017), sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). The total number of observations amounts to 4,142. Standard errors are clustered at the individual level. Whiskers represent 95% confidence intervals. *Source:* SILM Evaluation Dataset, see Brussig et al. (2019).

program more controls find a job. As a consequence, the average levels of well-being and social integration in the control group gradually increase and the ATT goes down.

Figure 3: Outcomes of participants, employed and unemployed control individuals



Notes: The figure shows conditional means of the outcome variables of participants, employed and unemployed control individuals at different program durations (mean duration of 7 months in t_1, 18 months in t_2, 29 months in t_3) based on the final estimation sample (see Section 4). Conditional means are predicted by OLS regressions of the standardized outcome variables on time indicators, their interactions with the treatment indicator and interaction terms of the time indicators and an employment indicator. Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). The total number of observations amounts to 7,550 (4,736 in t_1, 2,090 in t_2, 724 in t_3). The share of employed controls increases from 26% in t_1 to 32% in t_2 and 36% in t_3. Matches of drop-outs are excluded from the sample. Standard errors are clustered at the individual level. Whiskers represent 95% confidence intervals.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

These findings point to a trade-off with respect to the duration of the program. As there are no signs of habituation, a long program duration seems to be justified. However, the share of employed controls increases as the program proceeds. This suggests that many participants would have found a job if they had not been 'locked' into the program. Despite the program's focus on hard-to-place individuals, its effectiveness might be increased if access was further restricted to those individuals with the lowest job market prospects.

6.3 Robustness checks

In this subsection we provide robustness checks that relate to the selection problems introduced in Section 5 and potential misspecification.

(i) Job center and individual selection into the program. In a first robustness check, we exploit the fact that not all job centers implemented the program. Control individuals in non-participating job centers did not have the chance to sort into the program. Hence, this group might be more likely to represent a random draw from the population of potential controls. Control individuals of participating job centers, on the other hand, might have had the chance to enter the program, but did not do so. If, for example, control individuals with lower outcome levels are less likely to enter the program, our estimates would be biased upwards. Moreover, for individuals in participating job centers also spillover effects could bias the estimated treatment effects in an arbitrary direction. As outlined above, participating and non-participating job centers are likely to differ with respect to regional conditions and their strategies. A comparison of observable characteristics shows that control individuals of non-participating job centers are located in more prosperous labor market regions than controls of participating job centers.[25] We split the sample of matches in two subsets depending on whether or not the control individual is assigned to a participating job center. We then re-estimate the ATT for wave 1 within these subsamples, to check for any signs of selection in control group. Table B.5 in the Appendix shows that the effects on life satisfaction and the social integration measures are still significantly positive but smaller in magnitude for control individuals in participating job centers. The ATT on mental health is slightly higher. Except for life satisfaction, none of these differences is significant at the 10% level. As compared to the baseline results, the estimates indicate that selection on the job center and individual level may play a role to some extent but they provide no evidence against our main conclusions on the impact of the program.

In a further robustness check, we analyze whether selection into the program might be related to the availability of program places. Participants in job centers that offer less jobs per eligible persons might be a positive selection, as more potential participants compete for the same program place. Availability is measured by the ratio of program places over the number of welfare benefit recipients on the job center level. We split the sample into matches with participants at job centers with above and below average

[25]The results are available on request.

availability of program places (see Table B.6 in the Appendix).[26] The estimated ATTs in wave 1 are similar in both subsamples, providing no evidence for selection into the program. We come to the same conclusion when adding indicators of willingness to work as control variables (see Figure B.3 in the Appendix).[27]

(ii) Selection into the survey. Beside selection into the program, the results might be affected by non-random selection into the survey. Figure B.5 in the Appendix presents results for a specification weighted with the individual probability to respond to the first and follow-up surveys of program participants. This probability is estimated based on sociodemographic characteristics and the labor market history from the administrative records. For the follow-up waves, the survey data are used to control for lags of the outcome variables and a lagged indicator of program drop-out. Using the estimated response probabilities as weights in our baseline regression does not change the estimated effects. In a further check of non-random panel mortality, we re-estimate the ATTs using only observations from the balanced panel. The results are shown in Figure B.6 in the Appendix. As the sample size in each wave is smaller, the estimates from the balanced panel are more imprecise. They are also slightly lower than in the baseline model, but they are still in the same order of magnitude and show the same declining pattern over the waves. This suggests that the baseline estimates might be slightly upward biased by selective non-response to the survey. This bias, however, seems too small to change the main conclusions.

(iii) Empirical specification. By using a linear regression model, we follow a common practice in the literature although the ordinal outcome variables are thereby implicitly treated as if they were measured on a cardinal scale. Ferrer-i Carbonell and Frijters (2004), show that ordinal and cardinal estimation methods produce similar results when applied to ordinal measures of life satisfaction. One explanation might be that individuals tend to split up the response space into equally sized intervals when asked to assign a discrete number to verbal labels, e.g. ranging from "very bad" to "very good" (van Praag, 1991). Chen et al. (2019) argue, however, that treating ordinal variables as cardinal might render comparisons of group means invalid. Instead, regression models for ordered data should be applied and the results can be interpreted as comparisons in group medians. We obtain very similar results from ordered probit regressions (see Table B.7 in the Appendix).

In a further robustness check, we use a fixed-effects-model as an alternative to the pooled specification in equation 1 to identify the change in the ATTs between the

[26] The average job center offers about 2 places per 100 welfare benefit recipients.

[27] See Table A.2 for a description of the indicators of willingness to work and Figure B.4 for plots of the distributions.

waves.[28] The estimates from both the pooled and the differenced model are almost identical (see Table B.8 in the Appendix), which suggests that the negative time trends in the ATTs are not driven by unobserved individual fixed effects.

6.4 Heterogeneous effects and the role of program design

Employment not only provides material benefits but also satisfies psychological needs such as regular and meaningful activity, social contacts, participation in collective purpose and status and identity (Jahoda, 1981). Working in a JCS can substitute for these latent functions of work to some extent and thus improve the well-being and social integration of unemployed individuals. These effects might be larger for some groups and depend on specific program characteristics. This subsection digs deeper into potential mechanisms driving the main findings. First, we study effect heterogeneity with respect to sociodemographic characteristics and then analyze different aspects of the program design.

To explore effect heterogeneity, the estimation sample of matches is split along different participant characteristics and the program effect is estimated within these subsamples. Figure 4 plots the estimates for individuals with and without health impairments. The results point to a clear pattern: individuals with health impairments benefit more from program participation. The estimated ATTs in wave 1 are larger for each outcome variable. With respect to life satisfaction, the effects are also more stable over time than for individuals without health impairments.

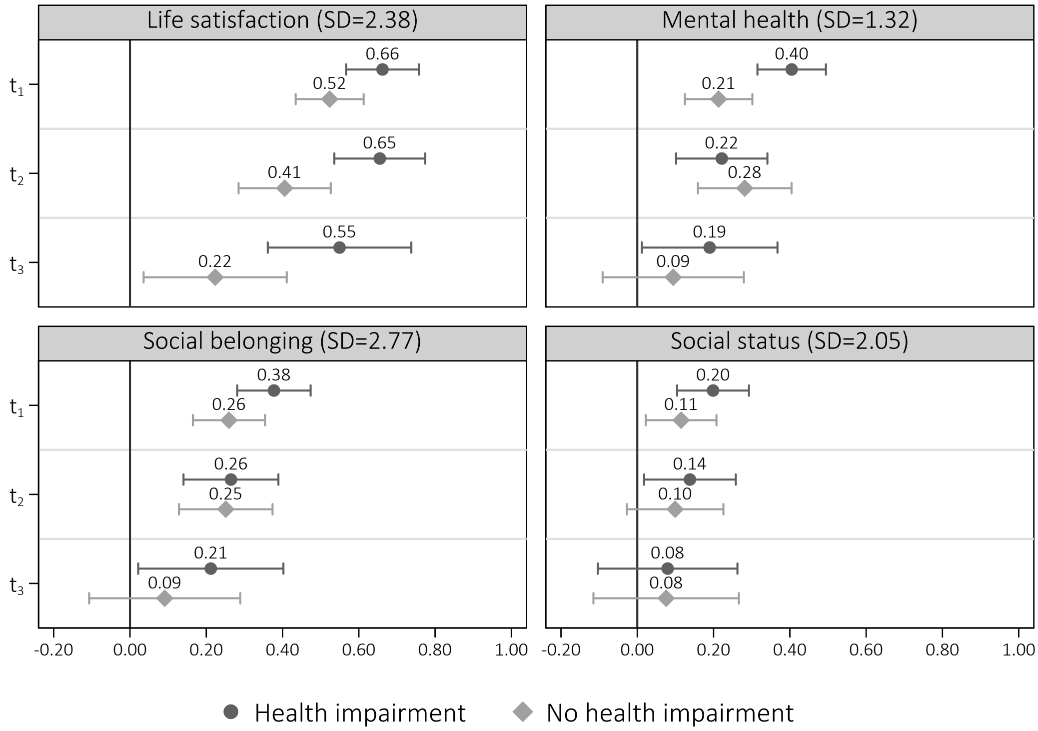
Figure 5 shows that the program is more beneficial for individuals receiving welfare for more than the sample mean of seven years prior to program entry. Effect heterogeneity is more pronounced for the social integration measures than for the well-being measures. While the effects for life satisfaction and mental health are decreasing over time for both groups, the estimated ATTs on social belonging and social status are more stable over time for participants with longer duration of welfare dependence. We also analyze effect heterogeneity with respect to other sociodemographic characteristics. Our findings suggest that having minor children, being female or prior participation in One-Euro-Jobs does not change the effect size significantly.[29]

Next, we turn to the role of different design aspects of SILM. The estimates pre-

[28]Note that this approach can only identify the change in the ATTs over the course of the program but not its level, because we do not have a pre-treatment measurement of the outcomes.

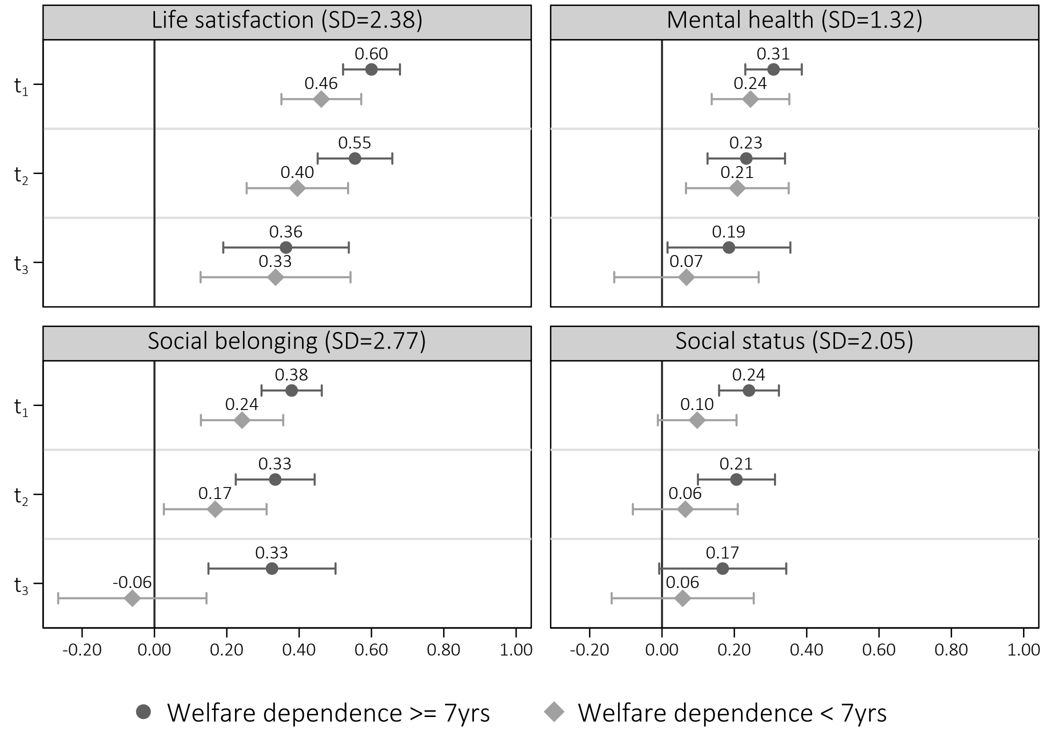
[29]We also study effect heterogeneity with respect to prior participation in the program 'Citizen Work', the need for support in four different life domains, marital status, age above 45 and for three education groups. We did not find significant differences in treatment effects for these groups. The results are available on request.

Figure 4: Estimated ATTs on standardized outcomes for participants with and without health impairments



Notes: SD: standard deviation. The figure shows the estimated ATTs at different program durations (mean duration of 7 months in t_1, 18 months in t_2, 29 months in t_3) from pooled OLS regressions (see equation 1) based on the final estimation sample (see Section 4). Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). The model is estimated separately for participants with and without health impairments and their matched control individuals. T-tests show that the estimated ATTs on life satisfaction always differ significantly (10% level) between sub-groups. For mental health and social belonging the difference is only significant in t_1. For social status, the estimated ATTs are not significantly different. The number of observations for individuals with health impairments amounts to 4,276 (2,544 in t_1, 1,258 in t_2, 474 in t_3). The number of observations for individuals without health impairments amounts to 4,010 (2,514 in t_1, 1,084 in t_2, 412 in t_3). Standard errors are clustered at the individual level. Whiskers represent 95% confidence intervals.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Figure 5: Estimated ATTs on standardized outcomes for participants with < and ≥ 7 years of prior welfare dependence



Notes: SD: standard deviation. The figure shows the estimated ATTs at different program durations (mean duration of 7 months in t_1, 18 months in t_2, 29 months in t_3) from pooled OLS regressions (see equation 1) based on the final estimation sample (see Section 4). Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). The model is estimated separately for participants with a prior welfare dependence duration below and above the sample mean of seven years and their matched control individuals. T-tests show that the estimated ATT on life satisfaction differs significantly (10% level) between sub-groups for t_1 and t_2. For social belonging, the difference is always significant. For social status, the estimated ATTs are significantly different only in t_1. For mental health, the estimated ATTs are not significantly different. The number of observations for individuals with above average welfare dependence amounts to 5,116 (3,114 in t_1, 1,472 in t_2, 530 in t_3). The number of observations for individuals with below average welfare dependence amounts to 3,170 (1,944 in t_1, 870 in t_2, 356 in t_3). Standard errors are clustered at the individual level. Whiskers represent 95% confidence intervals.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

sented in Table 3 are based on fixed effects regressions using the panel survey sample of participants (see step 4 in Section 4). The results show that entering an accompanying training measure alongside the program is associated with a 0.157 SDs higher social belonging and a 0.102 SDs higher social status. Supportive interventions by case workers or external coaches – e.g. to resolve conflicts at the work place or to give advice to employers – are weakly positively related to social status. Well-being and social integration decrease when participants drop-out of the program. This is particularly true for life satisfaction and social status. For those who quit the program due to health problems – which is the most common reason (see Table B.1 in the Appendix) – life satisfaction decreases by 0.289 SDs, social belonging by 0.280 SDs and social status by 0.197 SDs. There is no significant association between engaging in activities with other participants and well-being or social integration. However, these joint activities are associated with a 6 percentage point lower probability of program drop-out.[30] Supportive interventions by case workers or coaches are associated with a 5% lower risk of leaving the program early. Activities with others are related to a 6% lower probability to drop out. Although intensive coaching and additional activities with others are not directly related to higher outcomes, they may stabilize the employment relationship and prevent drop-outs and thereby indirectly promote social integration and well-being of participants. [31]

7 An assessment of costs and benefits

The analyses carried out so far suggest that SILM significantly increases individual well-being and social integration. But do these gains also outweigh the costs of the program? To answer this question, we approximate the effective costs of the program per participant and year and compare them to an estimate of the monetary equivalent of the utility gain of participants.[32]

The direct labor costs of the program consist of the labor income of participants and the employer share of social security contributions. According to our survey data, participants worked on average 28 hours per week. Given a flat rate of 18.9% for the employer share of social security contributions and a wage rate of 8.50 Euro per hour (8.84 Euro per hour after January 2017), the average wage costs amounted to about

[30] Selected coefficients from probit regressions for the drop-out probability are shown in Table B.9 in the Appendix.

[31] B.9 also shows that higher shares of previously unfinished ALMP measures are associated to a higher risk of drop-out. Moreover, individuals who report a need for support with in certain life domains like psychological problems or addiction have a higher risk of leaving the program early. The significant positive effects of the indicators of wave 2 and 3 reflect the growing share of drop-outs over time.

[32] For details of the cost benefit analysis see Appendix C.

Table 3: Effect of program characteristics on participant outcomes

	(1) Life satisfaction	(2) Mental health	(3) Social belonging	(4) Social status
Training/qualification measure	0.045	0.063*	0.157***	0.102**
	(0.033)	(0.033)	(0.038)	(0.040)
Activities with other participants	0.031	0.055	-0.037	-0.002
	(0.039)	(0.043)	(0.054)	(0.054)
Support by case workers/coaches	0.012	0.013	0.031	0.063*
	(0.030)	(0.032)	(0.034)	(0.035)
Drop-out	-0.196***	-0.045	-0.090	-0.107*
	(0.074)	(0.068)	(0.066)	(0.062)
Drop-out b/c of health problems	-0.289**	-0.146	-0.280***	-0.197*
	(0.120)	(0.109)	(0.103)	(0.102)
Observations	7,006 - 7,855	7,073 - 7,928	7,011 - 7,843	6,982 - 7,816

Notes: The table shows coefficient estimates of different program characteristics on standardized outcomes from fixed effects models based on survey data of participants (see Section 4). For each program characteristic, a separate model is estimated. The number of observations differs due to missing values. Standard errors are clustered at the individual level. ***/**/* mark significance at the 1/5/10 % level.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

14,700 Euros per job and year. However, the effective costs for the government were substantially lower, because participants paid income taxes and social security contributions and most of the remaining net income was deducted from their welfare claims. The average income after taxes was about 9,000 Euros per participant and year. As the welfare claims are not contained in our dataset, they are approximated by regional averages by household types using information from the survey data.[33] We find that in absence of the program participants would have received welfare payments of about 10,200 Euros per year. After accounting for income deduction and some of the additional costs (administration and counseling as well as training or qualification measures), the federal government's effective costs per participant and year are estimated to be about 4,600 Euros.

To put this figure into perspective, we use the life satisfaction approach to estimate the monetary equivalent of the utility gain from working in SILM (Frey and Stutzer, 2002; 2010). Life satisfaction can be viewed as the result of an individual's ability to convert resources into a good life. It thus serves as a proxy for individual utility which can be used to analyze the welfare implications of public policies. We follow the literature and add net household income from the survey as an explanatory variable to

[33] Welfare claims mainly depend on the number of household members in different age groups and regional housing costs. Therefore, the regional averages provided by the Statistical Office of the Federal Employment agency should be a reasonable approximation of individual welfare claims.

equation 1. The quotient of the estimated ATT and the coefficient of monthly household income provides a measure of the additional income that would be necessary to compensate an individual for not participating in the program.[34] The results in Table C.1 in the Appendix suggest that it would be necessary to provide an individual, on average, with a yearly net household income of about 27,700 Euros to achieve the same life satisfaction as from working in the program in wave 1. In waves 2 and 3, the compensating income amounts to about 25,400 and 19,100 Euros per year, respectively. The decreasing amount of money can be explained by the declining program effect on life satisfaction over time. The 95% confidence intervals span from about 10,700 to 35,500 Euros per year.

The underlying assumptions of this approach are that a linear equation is a useful approximation of the individual's utility from program participation and income and that the estimated income coefficient is unbiased. In addition, we do not observe all potential components of the cost function and it is assumed that the program does not crowd out existing jobs. Having these restrictions in mind, the results suggest that the individual utility gain of program participation is at least as high as the net costs of the government. In this sense the program can be rated as efficient.

8 Conclusion

Unemployment not only puts individuals under economic strain, it also deteriorates well-being, social integration and mental health of individuals. The consequences of unemployment might be especially pronounced for long-term welfare recipients with limited access to the regular labor market. Despite the size of this group, they are usually underrepresented in large scale surveys like SOEP and little is known about how active labor market policy affects their quality of life.

We address this gap by analyzing whether publicly subsidized employment is an efficient policy instrument to improve the well-being and social integration of the long-term unemployed. This study investigates a recent German job creation scheme that explicitly aimed at promoting social integration. The analysis is based on a combination of high quality social security data and a telephone survey of participants and matched control individuals. The results suggest that participation in the program significantly improves well-being and social integration. The effect sizes are largest for life satisfac-

[34] A similar approach has been used to value important life events like marriage, widowhood or job loss (see e.g. Clark and Oswald, 2002; Ferrer-i Carbonell and van Praag, 2002; Blanchflower and Oswald, 2004) and public bads such as air pollution (Luechinger, 2009), airport noise (Van Praag and Baarsma, 2005) or terrorism (Frey et al., 2009).

tion and smallest for social status. The program effects decrease over the course of the program. This decline is driven by a catch up in the control group of non-participants, rather than by a reduction of the outcomes among the treated. Given the relatively long program duration of up to three years, an increasing share of control individuals finds employment over time. As employed controls reach similar levels of well-being and social integration as program participants, the average treatment effects decrease. Our analyses thus provide no evidence for habituation or anticipation of a return into unemployment with the end of the program approaching. These findings have important implications for designing future programs: although participants benefit from working in the program, a considerable share of them might have found a job and reached the same levels of well-being and social integration as from working in the program. In order to increase the impact of such programs, access could be restricted to those individuals with severe employment impediments like long welfare dependence or health impairments. We find that these individuals also benefit more from participation.

As an innovative aspect, the program provides additional activities like training courses, supportive interventions by case workers or joint activities with other participants. Most of these activities are not directly related to well-being and social integration. However, they are related to a lower probability of dropping out of the program. This implies that accompanying activities can stabilize the employment relationship and thus improve the outcomes indirectly.

The assessment of costs and benefits suggests that the monetary equivalent of the individual utility gain from working in the program is at least as high as the net costs of the government. The net costs are relatively low because participants' earnings come along with increased tax revenues and savings in welfare payments.

Overall, our results suggest that targeted JCSs can be an efficient policy instrument to improve the social integration and well-being of the long-term unemployed. However, the program effects differ between the outcomes. Life satisfaction increases more than the feeling of social belonging and social status. The reasons for these differences need to be investigated in more depth in future research. This could yield valuable insights for the design and allocation of active labor market policies to improve the quality of life for long-term unemployed individuals who are not able to find a regular job.

References

Autor, D. H. and M. G. Duggan (2006). "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding". In: *Journal of Economic Perspectives* 20.3, pp. 71–96.

Autor, D. H., M. G. Duggan, K. Greenberg, and D. S. Lyle (2016). "The Impact of Disability Benefits on Labor Supply: Evidence from the VA's Disability Compensation Program". In: *American Economic Journal: Applied Economics* 8.3, pp. 31–68.

Berg, M., R. Cramer, C. Dickmann, R. Gilberg, B. Jesske, M. Kleudgen, A. Bethmann, B. Fuchs, M. Huber, S. Schwarz, M. Trappmann, and A. Reindl (2016). *Codebook and Documentation of the Panel Study 'Labour Market and Social Security'(PASS): Datenreport Wave 9. FDZ - Datenreport, 07/2016 (en)*. Tech. rep. infas Institute for Applied Social Science, Bonn and Institute for Employment Research (IAB), Nuremberg.

Bhuller, M., C. N. Brinch, and S. Königs (2017). "Time aggregation and state dependence in welfare receipt". In: *The Economic Journal* 127.604, pp. 1833–1873.

Blanchflower, D. G. and A. J. Oswald (2004). "Well-being over time in Britain and the USA". In: *Journal of Public Economics* 88.7-8, pp. 1359–1386.

Boockmann, B., T. Brändle, G. Klee, P. Kugler, K. Sippli, P. Fuchs, and H. Verbeek (2017). *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt – Zwischenbericht 2019, Forschungsbericht 533*. Bundesministerium für Arbeit und Soziales, Berlin.

Brussig, M., P. Aurich-Beerheide, J. Kirsch, P. Langer, A. Gabler, S. Kotlenga, B. Nägele, N. Pagels, B. Ivanov, F. Pfeiffer, L. Pohlan, R. Kleinemeier, and H. Puhe (2019). *Evaluation des Bundesprogramms ‚Soziale Teilhabe am Arbeitsmarkt' – Dritter Zwischenbericht, Forschungsbericht 531*. Bundesministerium für Arbeit und Soziales, Berlin.

Caliendo, M., R. Mahlstedt, and O. A. Mitnik (2017). "Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labor market policies". In: *Labour Economics* 46, pp. 14–25.

Card, D., J. Kluve, and A. Weber (2018). "What works? A meta analysis of recent active labor market program evaluations". In: *Journal of the European Economic Association* 16.3, pp. 894–931.

Chen, L., E. Oparina, N. Powdthavee, and S. Srisuma (2019). "Have Econometric Analyses of Happiness Data Been Futile? A Simple Truth About Happiness Scales". In: *IZA Discussion Paper* No. 12152.

Clark, A. E. (2003). “Unemployment as a social norm: Psychological evidence from panel data”. In: *Journal of Labor Economics* 21.2, pp. 323–351.

Clark, A. E. and A. Lepinteur (2019). “The Causes and Consequences of Early-Adult Unemployment: Evidence from Cohort Data”. In: *Journal of Economic Behavior & Organization* 166, pp. 107–124.

Clark, A. E. and A. J. Oswald (1994). “Unhappiness and unemployment”. In: *The Economic Journal* 104.424, pp. 648–659.

Clark, A. E. and A. J. Oswald (2002). “A simple statistical method for measuring how life events affect happiness”. In: *International Journal of Epidemiology* 31.6, pp. 1139–1144.

Crost, B. (2016). “Can workfare programs offset the negative effect of unemployment on subjective well-being?” In: *Economics Letters* 140, pp. 42 –47.

Dauth, W., M. Dorner, and U. Blien (2013). *Neukonzeption der Typisierung im SGB-II-Bereich*. Tech. rep. IAB-Forschungsbericht.

Di Tella, R., R. J. MacCulloch, and A. J. Oswald (2001). “Preferences over inflation and unemployment: Evidence from surveys of happiness”. In: *The American Economic Review* 91.1, pp. 335–341.

Ferrer-i Carbonell, A. and P. Frijters (2004). “How important is methodology for the estimates of the determinants of happiness?” In: *Economic Journal* 114.497, 641—659.

Ferrer-i Carbonell, A. and B. M.S. van Praag (2002). “The subjective costs of health losses due to chronic diseases. An alternative model for monetary appraisal”. In: *Health Economics* 11.8, pp. 709–722.

Frey, B. S. and A. Stutzer (2002). “What can economists learn from happiness research?” In: *Journal of Economic Literature* 40.2, pp. 402–435.

Frey, B. S. and A. Stutzer (2010). “Happiness and public choice”. In: *Public Choice* 144.3, pp. 557–573.

Frey, B. S., S. Luechinger, and A. Stutzer (2009). “The life satisfaction approach to valuing public goods: The case of terrorism”. In: *Public Choice* 138.3, pp. 317–345.

Gundert, S. and C. Hohendanner (2014). “Do fixed-term and temporary agency workers feel socially excluded? Labour market integration and social well-being in Germany”. In: *Acta Sociologica* 57.2, pp. 135–152.

Gundert, S. and C. Hohendanner (2015). “Active labour market policies and social integration in Germany: Do ‘One-Euro-Jobs’ improve individuals’ sense of social integration?” In: *European Sociological Review* 31.6, pp. 780–797.

Hetschko, C., A. Knabe, and R. Schöb (2014). “Changing identity: Retiring from unemployment”. In: *The Economic Journal* 124.575, pp. 149–166.

Huber, M., M. Lechner, and C. Wunsch (2011). “Does leaving welfare improve health? Evidence for Germany”. In: *Health Economics* 20.4, pp. 484–504.

Jahoda, M. (1981). “Work, employment, and unemployment: Values, theories, and approaches in social research.” In: *American Psychologist* 36.2, p. 184.

Kluve, J. (2010). “The effectiveness of European active labor market programs”. In: *Labour Economics* 17.6, pp. 904 –918.

Knabe, A., R. Schöb, and J. Weimann (2017). “The subjective well-being of workfare participants: insights from a day reconstruction survey”. In: *Applied Economics* 49.13, pp. 1311–1325.

Korpi, T. (1997). “Is utility related to employment status? Employment, unemployment, labor market policies and subjective well-being among Swedish youth”. In: *Labour Economics* 4.2, pp. 125 –147.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013). “Duration dependence and labor market conditions: Evidence from a field experiment”. In: *The Quarterly Journal of Economics* 128.3, pp. 1123–1167.

Luechinger, S. (Feb. 2009). “Valuing Air Quality Using the Life Satisfaction Approach”. In: *The Economic Journal* 119.536, pp. 482–515.

Paul, K. I. and K. Moser (2009). “Unemployment impairs mental health: Meta-analyses”. In: *Journal of Vocational behavior* 74.3, pp. 264–282.

Pohlan, L. (2019). “Unemployment and social exclusion”. In: *Journal of Economic Behavior & Organization* 164, pp. 273–299.

Trappmann, M., S. Gundert, C. Wenzig, and D. Gebhardt (2010). “PASS: A household panel survey for research on unemployment and poverty”. In: *Schmollers Jahrbuch: Zeitschrift für Wirtschafts- und Sozialwissenschaften* 130.4, pp. 609–622.

Van Praag, B. M. S. and B. E. Baarsma (2005). “Using Happiness Surveys to Value Intangibles: The Case of Airport Noise”. In: *The Economic Journal* 115.500, pp. 224–246.

Van Praag, B. M.S. (1991). “Ordinal and cardinal utility”. In: *Journal of Econometrics* 50, pp. 69–89.

Wulfgramm, M. (2011). “Can activating labour market policy offset the detrimental life satisfaction effect of unemployment?” In: *Socio-Economic Review* 9.3, pp. 477–501.

Appendix

A Data addendum

Table A.1: Description of variables based on Integrated Employment Biographies (IEB)

Variable	Description
Control variables measured at the date 31/12/14 (cohort 1) and 31/12/15 (cohort 2 and 3)	
Sociodemographics	
Female	Dummy for being female
Age	Dummies for age groups: 35 - 44 years, 45 - 54 years, $>$ 54 years, reference category is $<$ 35 years
Health impairment	Dummy for having serious health restrictions (includes officially recognized disabilities)
Minor children in household	Dummy for having children aged $\leq$ 18 in the household
Number of children in household	Number of children aged $\leq$ 15 in the household
German	Dummy for being German
Family status	Dummy for family status: married - separated, married, widowed and divorced, reference category is single
School degree	Dummies for highest school degree: *Sonderschulabschluss/Hauptschulabschluss*, *Mittlere Reife*, *Fachhochschulreife* and *Abitur*, reference category is no school degree
Professional degree	Dummies for highest professional degree: vocational training, *Abitur* only, *Abitur* and vocational training, academic degree of *Fachhochschule* and university degree, reference category is no vocational training
Previous job characteristics (for marginal employment and employment with ssc)	
Employment full-time	Dummy for being employed full-time
Job classifications	Dummies for 6 job classifications: 1 Farmer, 2 Production/Craftspeople/Technician, 3 White-collar employee, 4 Salesperson, 5 Clerical workers, 6 Service workers, reference category is 1
Tenure	Dummies for employment duration: categories are spitted according to percentiles of distribution: 25 - 50, 50 - 75, $>$ 75, reference category is 0 - 25
Daily wage	Dummies for daily wage in Euros (2010 prices): categories are spitted according to percentiles of distribution: 25 - 50, 50 - 75, $>$ 75, reference category is 0 - 25
Previous firm characteristics (for marginal employment and employment with ssc)	
Firm size	Dummies for number of employees: 10 - 49, 50 - 249, 250 - 499, $>$ 500, reference category is < 10

. . .

Table A.1: Description of variables based on IEB *(continuation)*

Variable	Description
Median wage of firm	Dummies for median wage in Euros (2010 prices): categories are spitted according to percentiles of distribution: 25 - 50, 50 - 75, $>$ 75, reference category is 0 - 25
Sector of firm	Dummies for 17 sectors: 1 Agriculture, 2 Goods production, 3 Metal , 4 Vehicles, 5 Consumption, 6 Food, 7 Construction I, 8 Construction II, 9 Wholesale, 10 Retail, 11 Transportation, 12 Services I, 13 Services II, 14 Education/Health, 15 Associations, 16 Public, 17 other, reference category is 1
Employment history	
Previous employment status	Dummies for previous employment status (multiple answers are possible): employed with social security contributions (ssc), marginally employed, unemployed, welfare claimant (UB II), unemployed with sick note, unemployed comment "difficult to place", non-employed (no data entry)
Previous participation in ALMP measures	Dummies for type of previous labor market policy measure: job creation scheme, employment subsidies, training, 'Citizen work' (*Bürgerarbeit*), 'One-Euro-Job' (*Arbeitsgelegenheiten*), 'Program at an employer' and 'Program at an institution' (*MAG and MAT*) and other.
Number of periods	Number of periods in respective labor market states
Duration of periods	Dummies for duration in respective labor market states: categories are spitted according to percentiles of distribution: 25 - 50, 50 - 75, $>$ 75, reference category is 0 - 25
SGB II-Types	Dummies indicating SGB II - Typ of job center: 1 - 15 (see Dauth et al., 2013). Job centers that face similar regional conditions are grouped into distinct types (so-called SGB-II-Types). Variables that are important determinants of these types are the density of jobs, the share of welfare claimants and the share of foreigners, among others. All in all, 15 SGB-II-Types characterize regions with an increasing share of SGB-II benefit claimants and decreasing placement prospects for the long-term unemployed.
Program and job characteristics	
Planned program duration	Program duration in months defined as difference between start and planned end
Welfare receipt	Dummy for being welfare claimant (UB II) during program duration
Daily wage	Daily wage in Euros (2010 prices)

Table A.2: Description of variables based on survey data

Variable	Description
Outcomes	
Life satisfaction	Categorical variable measuring life satisfaction ranging from 0 (completely dissatisfied) to 10 (completely satisfied)
Mental health	Categorical variable for assessment of mental health over the last 4 weeks ranging from from 1 (extreme problems) to 5 (no problems)
Social belonging	Categorical variable measuring perceived social affiliation ranging from 1 (feeling excluded) to 10 (feeling affiliated)
Social status	Categorical variable measuring assessment of position in society ranging from 1 (belonging to bottom) to 10 (belonging to the top)
Need for support	
Care of minor children	Dummy for need for support in the care of minor children
Psychological problems or addiction	Dummy for need for support with psychological problems or addiction
Indebtedness	Dummy for need for support with debts
Other life domains	Dummy for need for support in other life domains
Willingness to work	
Accept commute >1h	Dummy for willingness to accept (rather or absolutely) a commute of one hour of more to get a job (1 to 4 scale)
Accept unfavorable working hours	Dummy for willingness to accept (rather or absolutely) unfavorable working hours to get a job (1 to 4 scale)
Accept work below qualification level	Dummy for willingness to accept (rather or absolutely) tasks below the own qualification level to get a job (1 to 4 scale)
Accept unfavorable conditions	Dummy for willingness to accept (rather or absolutely) unfavorable conditions like noise or dirt to get a job (1 to 4 scale)
Accept employer with bad image	Dummy for willingness to accept (rather or absolutely) an employer with a bad image to get a job (1 to 4 scale)
Importance of earning own money	Dummy for agree (rather of fully) with statement "Earning my own money is important to me" (1 to 5 scale)
Program and job characteristics	
Program drop-out	Dummy for program drop-out
Employment-accompanying activities	Dummies for employment-accompanying activities: personal counseling, training/qualification measure, Support by case worker/coach, activities with other participants, health promotion
Average working hours per week	Average working hours per week

...

Table A.2: Description of variables based on survey data *(continuation)*

Variable	Description
Program tasks	Dummies for program tasks: social work, gardening/crafts/janitor, administration/archive/library, kitchen/food distribution, cleaning/housekeeping, sales/social department stores, others
Participants assessments of the program	
Job satisfaction	Categorical variable measuring job satisfaction ranging from 0 (completely dissatisfied) to 10 (completely satisfied)
Work is meaningful	Dummy for agree (rather or fully) with statement "I perceive my work as meaningful" (1 to 4 scale)
Good relations with colleagues	Dummy taking the value of one if the average level of approval to the following four items is greater or equal to 3 (1 to 4 scale): "I receive help and support from my colleagues if needed", "Overall, I am treated fairly at the workplace", "I am acknowledged by my superior(s)", "At work I am more or less on my own" (scale flipped such that higher value marks less approval).

Table A.3: Construction of the SILM Evaluation Dataset

Step	Description	Observations
Step 1	Identification of program participants in administrative data.	12,412 participants
Step 2	Pre-selection of 20 control individuals for each program participant along key characteristics in administrative data.	248,240 controls
Step 3	Nearest-neighbor propensity score matching with replacement to select 4 control individuals for each participant.	12,412 participants, 49,648 controls
Step 4	*Telephone survey* of all program participants and matched control individuals.	
Step 4.1	*Wave 1:* Survey of all program participants as soon as possible (average time since entry: 7 months) and of one nearest neighbor.	3,821 participants, 3,451 controls
Step 4.2	*Wave 2*: Survey of all participants (average time since entry: 18 months) and controls from wave 1.	2,711 participants, 2,189 controls
Step 4.3	*Wave 3:* Survey of all participants (average time since entry: 29 months) and controls from wave 2. No third wave for late entrants (after 01/2017).	1,415 participants, 1,135 controls
Step 5	*Final sample of matched pairs:* Restriction to successfully surveyed matched pairs with non-missing information on key variables. Exclude control individuals in alternative treatment ("ESF-federal program", see Boockmann et al., 2017).	
	Wave 1 (t_1):	2,529 participants, 2,529 controls
	Wave 2 (t_2):	1,171 participants, 1,171 controls
	Wave 3 (t_3):	443 participants, 443 controls

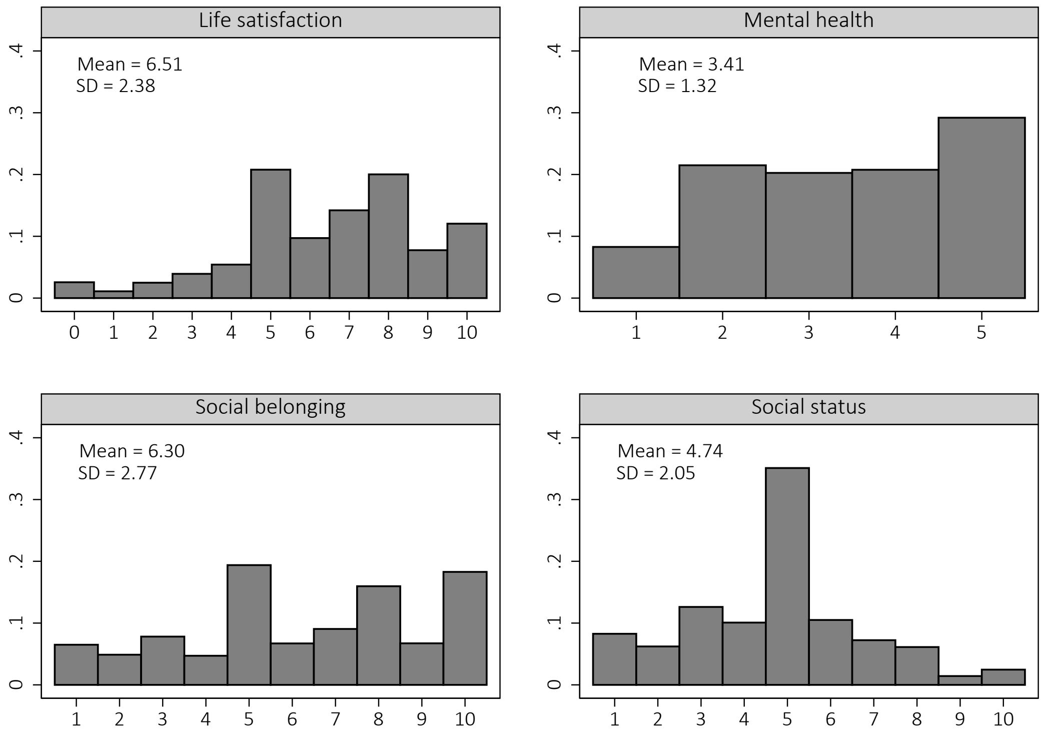
B Descriptive statistics and further results

Table B.1: Reason for dropping out of the program and employment status after drop-out, mean values by survey wave

	Mean		
	Wave 1	Wave 2	Wave 3
Reasons for dropping out:			
Health reasons/physical or mental overload	0.494	0.423	0.325
Task not as expected/conflicts at workplace	0.286	0.266	0.205
Laid off by employer/job was terminated	0.104	0.100	0.084
Employment outside the program	0.071	0.154	0.361
Employment status after drop-out:[§]			
Unemployed	0.751	0.651	0.422
Full-time employment	0.083	0.141	0.241
Part-time or marginal employment	0.083	0.100	0.205
Other	0.154	0.261	0.289
Well-being and social integration:			
Life satisfaction	5.895	5.967	6.699
Mental health	3.021	3.116	3.373
Social belonging	6.060	6.218	6.338
Social status	4.417	4.376	4.321
Observations	241	241	83

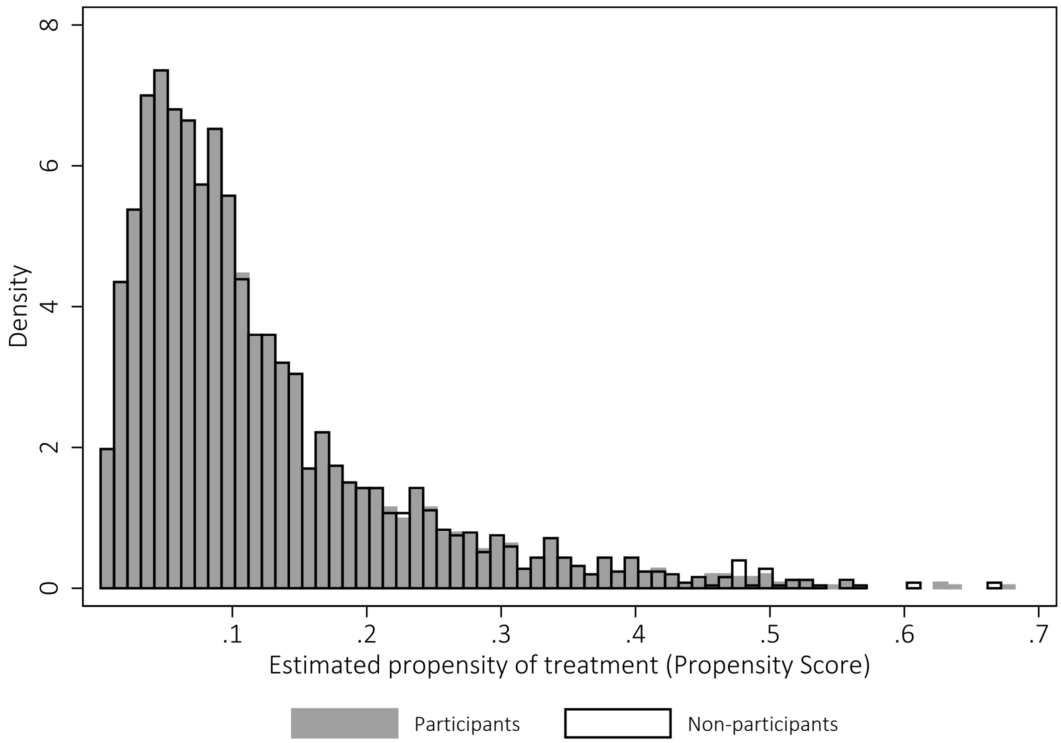
Notes: The table shows the means of indicators of different reasons to drop out of the program and the employment status afterwards based on survey data of participants (see Section 4). [§] Multiple answers possible. Other employment statuses contain e.g. subsidized employment, training, disability, retirement or sick leave.
Source: SILM Evaluation Dataset (see Brussig et al., 2019).

Figure B.1: Distribution of outcomes variables



Notes: SD: standard deviation. The figure shows the densities of the outcome variables in their original scale and provides means and standard deviations in the final estimation sample pooled over waves 1 to 3 (see Section 4).
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Figure B.2: Common support of estimated propensity of treatment (propensity score)



Notes: The figure shows the estimated propensity of participating in the program based on the final estimation sample (see Section 4).
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Table B.2: Comparison of participants throughout the sample construction process

	(1)	(2)	(3)
Step of sample construction process:	Step 1	Step 4.1	Step 5
Sample:	(Administrative data)	(Survey sample, wave 1)	(Final sample of matched pairs, wave 1)
Socio-demographics			
Female	0.43	0.47	0.47
Age [years]	48.39	48.85	48.92
Health impairment	0.47	0.49	0.50
Minor children in household	0.26	0.28	0.27
German	0.91	0.92	0.93
Married	0.28	0.25	0.25
No professional degree	0.20	0.17	0.16
Vocational degree	0.76	0.73	0.73
Acadamic degree	0.04	0.04	0.04
Region with weak placement prospects §	0.75	0.78	0.79
Employment history			
Years of welfare receipt	7.50	7.48	7.43
Cum. years of ssc. employment	6.09	6.47	6.53
Cum. years of marginal employment	1.06	1.29	1.27
Cum. number of ALMP measures	6.20	6.51	6.65
Prior participation in "One-Euro-Job"	0.71	0.74	0.77
Share of unfinished ALMP measures †	0.12	0.12	0.12
Need for support			
Care of minor children	-	0.10	0.09
Psychological problems or addiction	-	0.15	0.15
Indebtedness	-	0.19	0.20
Other life domains	-	0.21	0.21
At least one field	-	0.45	0.44
Outcomes			
Life satisfaction [0-10]	-	7.18	7.14
Mental health [1-5]	-	3.66	3.64
Social belonging [1-10]	-	6.90	6.87
Social status [1-10]	-	5.03	4.98
Observations	12,412	3,535	2,529

Notes: The table shows the means of selected participant characteristics in different stages of the sampling process (see Section 4). Column (1) shows means from the initial sample of participants drawn from the administrative records. Column (2) shows means from the sample participants who entered the survey in wave 1. The sample is restricted to participants with non-missing values for all variables. Column (3) shows the mean values in wave 1 of the final estimation sample. Sociodemographics and the employment history are based on administrative data. For details see Table A.1 and A.2 in the Appendix. § Defined as job center regions of SGB II-Type 3 according to the classification of the Federal Employment Agency (Dauth et al., 2013). † An ALMP spell is defined as unfinished if it ends before the planned end date without a transition into employment.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Table B.3: Comparison of outcomes in SILM Evaluation Dataset and PASS

	(1)	(2)	(3)	(4)
	Final sample of matches		PASS	
	Participants	Non-participants	Employed	Eligible
Well-being				
Life satisfaction [0-10]	7.14	5.73	7.30	5.35
Mental health [1-5]	3.64	3.22	4.00	3.36
Social integration				
Social belonging [1-10]	6.87	5.88	7.80	5.60
Social status [1-10]	4.98	4.57	6.10	4.47
Observations	2,529	2,529	16,600	4,600

Notes: The table shows means of the outcome variables for participants (column (1)) and matched non-participants (column (2)) in the final estimation sample (see Section 4). Column (3) shows the mean outcomes for a sample of employed individuals and column (4) shows mean outcomes for individuals who fulfill the eligibility criteria of SILM based on PASS.
Source: SILM Evaluation Dataset, see Brussig et al. (2019); PASS, see Trappmann et al. (2010).

Table B.4: Balancing in pre-treatment characteristics and post-treatment outcomes of participants and controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Wave 1			Wave 2			Wave 3	
	participants	non-participants	Diff	participants	non-participants	Diff	participants	non-participants	Diff
Sociodemographics									
Female	0.47	0.45	0.02	0.48	0.46	0.01	0.48	0.47	0.01
Age [years]	48.92	48.77	0.15	49.86	49.23	0.64 **	50.05	49.18	0.86 *
Health impairment	0.50	0.48	0.02	0.54	0.49	0.05 **	0.53	0.50	0.03
Minor children in household	0.27	0.28	-0.01	0.25	0.26	-0.02	0.23	0.25	-0.01
German	0.93	0.93	-0.01	0.94	0.94	0.00	0.96	0.93	0.03 **
Married	0.25	0.25	0.00	0.27	0.25	0.02	0.25	0.22	0.03
No professional degree	0.16	0.16	0.00	0.15	0.15	0.00	0.16	0.15	0.01
Vocational degree	0.73	0.74	-0.01	0.75	0.75	0.00	0.74	0.74	0.00
Acadamic degree	0.04	0.04	0.01	0.05	0.05	0.01	0.06	0.05	0.01
Region with weak placement prospects §	0.79	0.77	0.01	0.79	0.77	0.02	0.77	0.77	0.00
Employment history									
Years of welfare receipt	7.43	7.21	0.22 **	7.49	7.29	0.20	7.14	6.96	0.18
Cum. years of ssc. employment	6.53	6.68	-0.15	6.84	6.82	0.02	7.00	7.14	-0.14
Cum. years of marginal employment	1.27	1.35	-0.07	1.30	1.51	-0.21 *	1.36	1.39	-0.03
Cum. number of ALMP measures	6.65	6.81	-0.16	6.62	6.81	-0.19	6.90	7.15	-0.26
Share of unfinished ALMP measures †	0.77	0.75	0.01	0.78	0.76	0.02	0.78	0.79	-0.01
Prior participation in "One-Euro-Job"	0.12	0.12	0.01	0.12	0.11	0.01	0.11	0.11	0.00
Need for support									
Care of minor children	0.09	0.07	0.02 ***	0.09	0.07	0.02 *	0.09	0.06	0.03 *
Psychological problems or addiction	0.15	0.16	-0.01	0.15	0.17	-0.01	0.12	0.18	-0.05 **
Indebtedness	0.20	0.17	0.02 **	0.18	0.16	0.01	0.18	0.16	0.02
Other life domains	0.21	0.28	-0.07 ***	0.21	0.28	-0.06 ***	0.20	0.28	-0.08 ***
At least one field	0.44	0.48	-0.03 **	0.44	0.47	-0.03	0.42	0.49	-0.07 *
Outcomes									
Life satisfaction [0-10]	7.14	5.73	1.41 ***	7.31	6.06	1.25 ***	7.00	6.04	0.96 ***
Mental health [1-5]	3.64	3.22	0.43 ***	3.57	3.22	0.34 ***	3.50	3.21	0.29 ***
Social belonging [1-10]	6.87	5.88	0.99 ***	6.67	5.85	0.81 ***	6.28	5.68	0.60 ***
Social status [1-10]	4.98	4.57	0.41 ***	4.89	4.57	0.32 ***	4.72	4.44	0.28 **
Observations	2,529	2,529		1,171	1,171		443	443	

Notes: The table shows means of selected pre-treatment characteristics of participants and matched non-participants in waves 1 to 3 of the final estimation sample (see Section 4). Column (3), (6) and (9) show the differences in means and their significance levels from two-sample t-tests. Sociodemographics and the employment history are based on administrative data. The questions regarding the need for support are measured in survey wave 1. For details see Table A.1 and A.2 in the Appendix. § Defined as job center regions of SGB II-Type 3 according to the classification of the Federal Employment Agency (Dauth et al., 2013). † An ALMP spell is defined as unfinished if it ends before the planned end date without a transition into employment. ***/**/* marks significance at the 1/5/10% level.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Table B.5: Estimated ATTs on standardized outcomes by job center type of matched control individuals

	(1)	(2)	(3)
	JC type of matched control		
	participating	non-participating	Difference
Life satisfaction	0.523***	0.629***	-0.106 *
	(0.033)	(0.046)	
Mental health	0.335***	0.259***	0.076
	(0.033)	(0.046)	
Social belonging	0.306***	0.387***	-0.081
	(0.036)	(0.048)	
Social status	0.134***	0.231***	-0.097
	(0.036)	(0.048)	
Observations	3,276	1,782	

Notes: The table shows the estimated ATTs in t_1 (mean duration of 7 months) from pooled OLS regressions (see equation 1) based on the final estimation sample (see Section 4). Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). Estimates are based on the subsample of matches with control individuals from job centers participating in the program (column (1)) and non-participating job centers (column (2)). Column (3) shows the difference between column (1) and (2) and its significance from two sample t-tests. Standard errors are clustered at the individual level. ***/**/* marks significance at the 1/5/10% level.

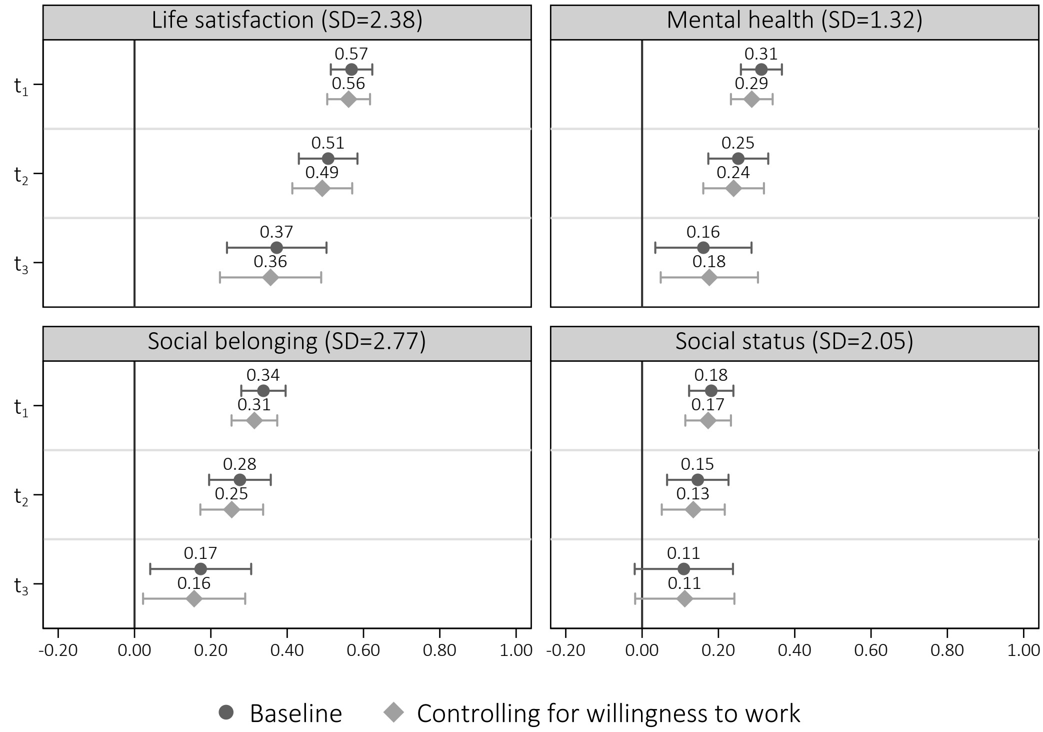
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Table B.6: Estimated ATTs on standardized outcomes by availability of program places in the job centers of participants

	(1)	(2)	(3)
	Availability of program places in participant JC		
	Above avrg.	Below avrg.	Difference
Life satisfaction	0.565***	0.566***	-0.001
	(0.047)	(0.034)	
Mental health	0.274***	0.319***	-0.045
	(0.045)	(0.034)	
Social belonging	0.379***	0.327***	0.052
	(0.048)	(0.036)	
Social status	0.228***	0.175***	0.053
	(0.049)	(0.036)	
Observations	1,926	3,132	

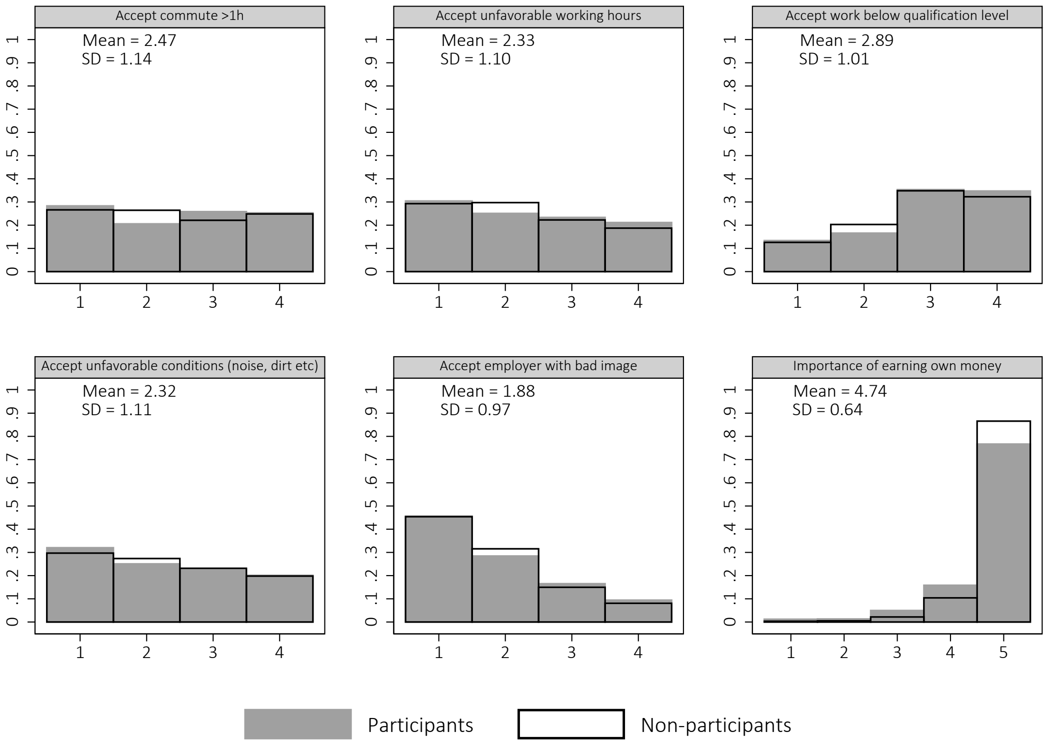
Notes: The table shows the estimated ATTs in t_1 (mean duration of 7 months) from pooled OLS regressions (see equation 1) based on the final estimation sample (see Section 4). Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). Estimates are based on the subsample of matches with treated individuals from job centers with above (column (1)) and below average availability of program places (column (2)). Availability is approximated by the number of program places divided by the number of welfare recipients per job center. The average availability amounts to 0.022, i.e. 2.2 program places per 100 welfare recipients. Column (3) shows the difference between column (1) and (2) and its significance from two sample t-tests. Standard errors are clustered at the individual level. ***/**/* marks significance at the 1/5/10% level.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Figure B.3: Estimated ATTs on standardized outcomes, with and without additional control variables of willingness to work



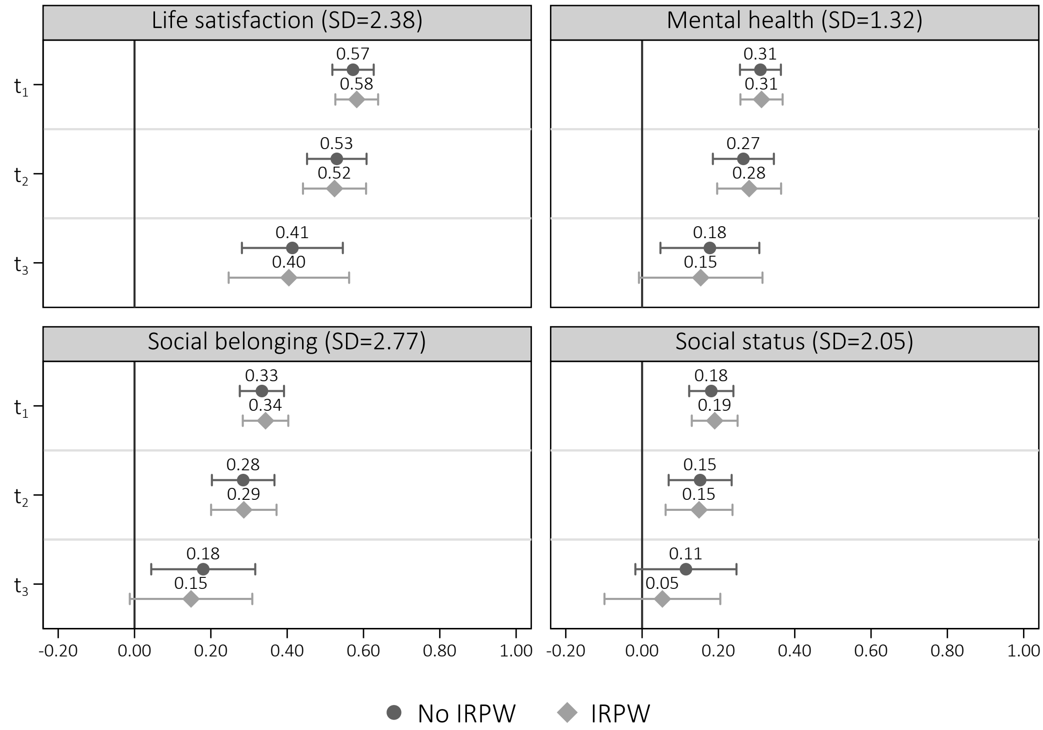
Notes: The figure shows the estimated ATTs at different program durations (mean duration of 7 months in t_1, 18 months in t_2, 29 months in t_3) from pooled OLS regressions (see equation 1) based on the final estimation sample (see Section 4). The model is estimated with and without additional control variables for willingness to work. Further control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). The variables reflecting willingness to work are derived from survey data and contain indicators of individual willingness to accept different adverse working conditions like noise, dirt or longer commutes in order to get a job and the importance of earning own money (see Table A.2). For the baseline model, the total number of observations amounts to 8,138 (5,052 in t_1, 2,258 in t_2, 828 in t_3). Due to missing values the total observation number for the model with controlling for willingness to work amounts to 6,972 (4,364 in t_1, 1,900 in t_2, 708 in t_3). Standard errors are clustered at the individual level. Whiskers represent 95% confidence intervals.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Figure B.4: Distribution of indicators of willingness to work for participants and non-participants



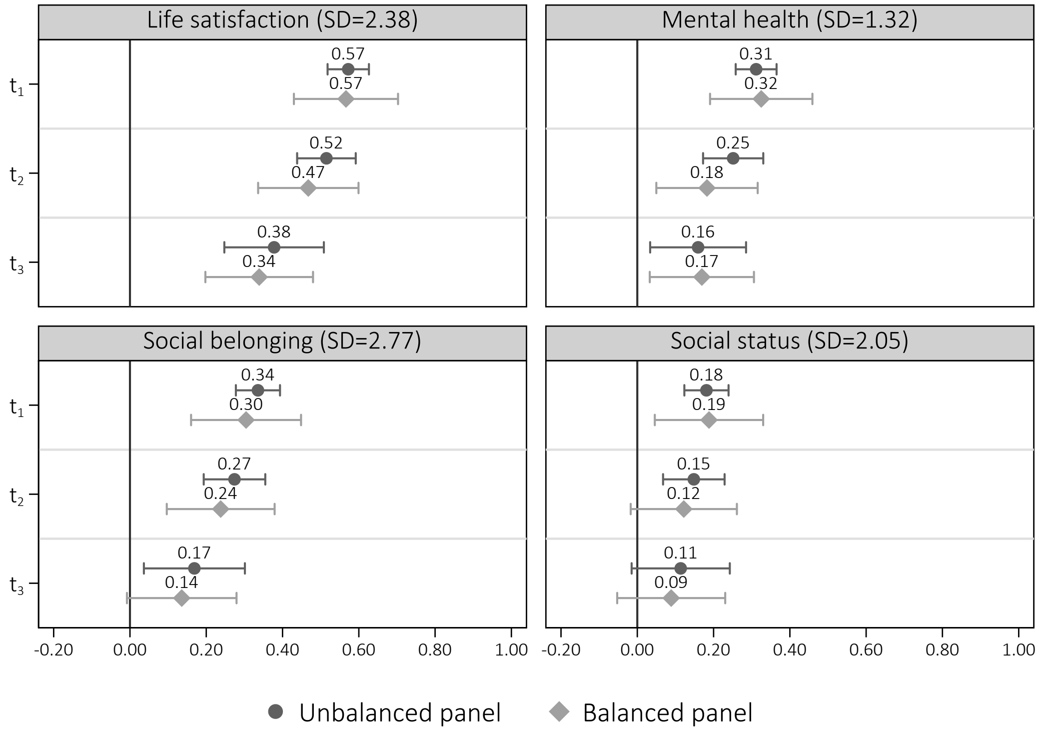
Notes: SD: standard deviation. The figure shows the densities of the control variables for 'work identity' (see Table A.2) separately for participants and non-participants in the final estimation sample pooled over waves 1 to 3 (see Section 4). Means and standard deviations are computed for the full sample (participants and non-participants).
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Figure B.5: Estimated ATTs on standardized outcomes, with and without inverse response probability weighting (IRPW)



Notes: IRPW: inverse response probability weighting, SD: standard deviation. The figure shows the estimated ATTs at different program durations (mean duration of 7 months in t_1, 18 months in t_2, 29 months in t_3) from pooled OLS regressions (see equation 1) based on the final estimation sample (see Section 4). Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). The model is estimated with and without IRPW. The probability to respond to wave 1 of the survey is estimated based on pre-treatment characteristics from the administrative data only. The probability of responding to the follow-up surveys is estimated based on additional information from survey data such as information on the need for support, a drop-out indicator and lags of the outcome variables and job satisfaction. The total number of observations amounts to 8,138 (5,052 in t_1, 2,258 in t_2, 828 in t_3). The sample is restricted to observations with non-missing estimated response probability. Standard errors are clustered at the individual level. Whiskers represent 95% confidence intervals.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Figure B.6: Estimated ATTs on standardized outcomes, estimated on unbalanced and balanced panel



Notes: SD: standard deviation. The figure shows the estimated ATTs at different program durations (mean duration of 7 months in t_1, 18 months in t_2, 29 months in t_3) from pooled OLS regressions (see equation 1) based on the final estimation sample (see Section 4). Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). The model is estimated in the unbalanced and balanced panel. The total number of observations in the unbalanced panel amounts to 8,286 (5,058 in t_1, 2,342 in t_2, 886 in t_3). For the balanced panel, the sample is restricted to matches with non-missing values in outcomes and control variables. This results in a total of 2,352 observations (784 in each wave). Whiskers represent 95% confidence intervals.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Table B.7: Estimated ATTs on standardized outcomes, estimates from ordered probit models

	(1)	(2)	(3)	(4)
	Well-being		Social integration	
	Life satisfaction	Mental Health	Social belonging	Social status
t_1	0.573***	0.303***	0.331***	0.180***
	(0.027)	(0.027)	(0.029)	(0.029)
t_2	0.506***	0.256***	0.266***	0.158***
	(0.039)	(0.040)	(0.041)	(0.041)
t_3	0.341***	0.177***	0.155**	0.116*
	(0.065)	(0.065)	(0.068)	(0.065)

Notes: The table shows the estimated ATTs at different program durations (mean duration of 7 months in t_1, 18 months in t_2, 29 months in t_3) from ordered probit models based on the final estimation sample (see Section 4). Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). For each wave, a separate model was estimated in the cross section of matched participants and controls. The total number of observations amounts to 8,286 (5,058 in t_1, 2,342 in t_2, 886 in t_3). Standard errors are clustered at the individual level. ***/**/* marks significance at the 1/5/10% level.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Table B.8: Estimated ATTs on standardized outcomes, estimates from a pooled and a differenced model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Life Satisfaction		Mental health		Social beloning		Social status	
	Pooled	Differenced	Pooled	Differenced	Pooled	Differenced	Pooled	Differenced
$Treat$	0.572***	-	0.312***	-	0.336***	-	0.181***	-
	(0.028)		(0.027)		(0.030)		(0.030)	
$Treat \times t_2$	-0.057	-0.081**	-0.060	-0.079**	-0.061	-0.076*	-0.033	-0.062
	(0.041)	(0.040)	(0.041)	(0.039)	(0.045)	(0.046)	(0.046)	(0.049)
$Treat \times t_3$	-0.194***	-0.227***	-0.152**	-0.157**	-0.167**	-0.147**	-0.067	-0.106
	(0.067)	(0.068)	(0.065)	(0.066)	(0.069)	(0.075)	(0.068)	(0.074)

Notes: The table shows the estimated ATTs in t_1 and the changes in t_2 and t_3 based on the final estimation sample (see Section 4). The pooled models (columns (1), (3), (5) and (7)) use the following specification to predict the levels of the outcomes variables: $y_{it} = \alpha_0 + \beta_1 Treat_i + \beta_2 Treat_i \times t_{2,it} + \beta_3 Treat_i \times t_{3,it} + \alpha_1 t_{2,it} + \alpha_2 t_{3,it} + X_{it}\gamma + c_i + \varepsilon_{it}$. The coefficients of the interactions terms are the change of the treatment effect as compared to t_1. c_i is a time-invariant and unobserved individual-specific error term. The estimates from this model can be compared to those from a "differenced" model (columns (2), (4), (6) and (8)) that use the change in outcomes w.r.t. t_1 as the dependent variable: $\Delta y_{it} = y_{it} - y_{i1} = \delta_0 + \beta_2 Treat_i \times t_{2,it} + \beta_3 Treat_i \times t_{3,it} + \delta_1 t_{2,it} + \delta_2 t_{3,it} + X_{it}\theta + \eta_{it}$. Control variables contains sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). The total number of observations amounts to 8,286 (5,058 in t_1, 2,342 in t_2, 886 in t_3). Standard errors are clustered at the individual level. ***/**/* marks significance at the 1/5/10% level.

Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Table B.9: Selected predictors of drop-out probability

	Probability of drop-out
Wave 2	0.047***
	(0.007)
Wave 3	0.099***
	(0.011)
Share of unfinished ALMP measures	0.072**
	(0.029)
Need for support:	
Psychological problems or addiction	0.036***
	(0.011)
Other life domains	0.037***
	(0.010)
Employment-accompanying activities:	
Training/qualification measure	-0.020*
	(0.011)
Support by case worker/coach	-0.048***
	(0.010)
Activities with other participants	-0.062***
	(0.017)
Observations	7,669

Notes: The table shows the average marginal effects of selected predictors of the probability of program drop-out based on survey data of participants (see Section 4). Estimates are obtained from a probit model of switching to the state of being a drop-out with indicators of survey waves and for employment-accompanying activities. Control variables contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). Standard errors are clustered at the individual level. ***/**/* mark significance at the 1/5/10 % level.
Source: SILM Evaluation Dataset (see Brussig et al., 2019).

C Details of cost benefit analysis

C.1 Approximation of costs

The effective costs of the program consist of two components: (i) the effective labor costs and (ii) additional costs for administration and employment-accompanying activities. We do not observe these cost components directly, so we combine institutional information, survey data and regional averages to approximate them. This subsection provides the details of this analysis.

(i) Effective labor costs. The direct labor costs of the program consist of the labor income of participants and the employer share of social security contributions. However, participants pay income taxes and social security contributions on their gross income and most of the net income is then deducted from their welfare claims. Therefore, most of the income from working in the program substitutes the welfare payments participants would receive otherwise. Since we do not observe the welfare claims and earnings of participants, we approximate them in the following way:

In Germany, social welfare benefits are means tested on the household level. The basic costs of living that are covered by the welfare system mainly consist of a fixed lump sum for each household member and a subsidy for the costs of accommodation which depend on the size and composition of the household (additional components are, for example, subsidized health insurance or payments for exceptional expenditures). These costs mainly vary by household size and housing costs. We approximate them using statistics of the Federal Employment Agency. These data contain the average values of the basic costs of living for different household types in different job center regions. In particular, the averages are available for single households, single parent households and cohabiting households with and without children. These information are matched to the survey data.

If participants are single earners, it is assumed that their income from working in the program is the only source of household income. In addition, child care benefits are considered, but no other sources of income. In case participants cohabit with an employed partner,the partner's income is approximated by the average of net earnings of employed individuals living in cohabiting households that are eligible to welfare payments in the PASS (N=1,440). To calculate the net income, income taxes are subtracted according to the German Income Tax Act (see Article 38) and a flat of 18.9% for social security contributions.

The deductible income is calculated from the gross income based on the following exemption rules (see Article 11b(2) SGB II): the first 100 Euros are completely exempt, for any earnings between 100.01 Euros and 1,000 Euros per month 20% are exempt, for any earnings between 1,000.01 and 1,200 Euros an additional amount of 10% are exempt from deduction.

The households' welfare claims are computed as the difference between the basic costs of living and deductible income. To calculate the effective costs of the program, the net income of program participants is compared to the welfare payments they would have received in absence of the treatment. The effective labor costs are computed as the difference between the participants' net household income and the welfare claims as provided by the above mentioned approximations. The resulting approximation amounts to about 3,350 Euros per participant and year.

(ii) Additional costs for administration and employment-accompanying activities. There are no explicit measures for these cost components available to us. By law, the common ratio of case workers and customers is 1:150 (for individuals above 25 years, see § 44c Abs.4 SGB II). Gross earnings of the average case worker amount to 49,020 Euro per year and full-time equivalent (income group E9 b TVöd level 5). This results in personal costs of about 327 Euros per year and customer. According to the survey data, participants have about twice as many appointments with their case workers than the control individuals. Assuming a ratio of 2:150, the additional personal costs for administration and counseling amount to 327 Euros per year and participant. Moreover, we take into account costs of the Federal Administration Office (*Bundesverwaltungsamt*) for program administration amounting to 21 per person and year.

To proxy some of the additional costs due to employment-accompanying activities, we use the share of individuals in training or qualification measures as provided by our survey. According to the budgetary statistics of Federal Employment Agency (*'Eingliederungsbilanzen'*) the average training measure costs 5,386.50 Euro per head.

The total additional costs per participant and year are calculated as the sum of the effective labor costs (after accounting for taxes, social security contributions and welfare savings), the additional administrative costs and the additional costs for training and qualification measures. They amount to about 4,600 Euro per year and participant. All other costs are disregarded.

Table C.1: Estimated compensating income

	(1)	(2)	(3)
	Coefficient estimates		Compensating income
	Life satisfaction	Life satisfaction	in 1,000 Euros/year
ATT:			
t_1	0.580***	0.541***	27.761
	(0.029)	(0.030)	(20.453 - 35.070)
t_2	0.539***	0.496***	25.446
	(0.042)	(0.042)	(17.933 - 32.960)
t_3	0.414***	0.372***	19.087
	(0.070)	(0.069)	(10.734 - 27.440)
Net household income	-	0.019***	
		(0.002)	
Observations	7,287	7,287	

Notes: The table shows estimates of the ATT on life satisfaction over the course of the program excluding yearly net household income (column (1)) and including yearly net household income (column (2)) based on the final estimation sample (see Section 4) with non-missing information on net household income. Control characteristics contain sociodemographics, the individual employment history and the need for support (see Tables A.1 and A.2 in the Appendix). Net household income is measured in 1,000 Euros per year. Standard errors are clustered at the individual level and shown in parentheses. ***/**/* marks significance at the 1/5/10% level. Column (3) shows estimates of the compensating income (CI) of participating in the program for the three survey waves and 95% confidence intervals in parentheses. The CI is the ratio of the estimated ATT and the estimated coefficient of household income.
Source: SILM Evaluation Dataset, see Brussig et al. (2019).

Download ZEW Discussion Papers from our ftp server:

http://ftp.zew.de/pub/zew-docs/dp/

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

//
IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de