I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17347

Feeling Observed? A Field Experiment on the Effects of Intense Survey Participation on Job Seekers' Labour Market Outcomes

Gesine Stephan
Clemens Hetschko
Julia Schmidtke
Michael Eid
Mario Lawes

OCTOBER 2024

IZA – Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17347

Feeling Observed? A Field Experiment on the Effects of Intense Survey Participation on Job Seekers' Labour Market Outcomes

Gesine Stephan
IAB, FAU and IZA

Clemens Hetschko
University of Leeds, FU Berlin and CESifo

Julia Schmidtke
IAB

Michael Eid
FU Berlin

Mario Lawes
FU Berlin

OCTOBER 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Feeling Observed? A Field Experiment on the Effects of Intense Survey Participation on Job Seekers' Labour Market Outcomes*

We ran a field experiment to causally identify the effects of intense survey participation on key labour market outcomes. We randomly excluded individuals willing to sign up for the German Job Search Panel, a high-frequency survey with a focus on job search and well-being. Using administrative data on labour market outcomes (e.g., employment, earnings), we find that, on average, survey participation had no effect on labour market outcomes during the year after signing up. Furthermore, there is no strong heterogeneity across subgroups. Overall, this is good news for the validity of survey-based research involving labour market outcomes. We also demonstrate that a comparison of individuals signing up for the survey with individuals not responding to the invitation could have been misleading. Even when controlling for a wide range of observable characteristics, survey participation and the subsequent take up of training programs correlate significantly. This speaks to the importance of experimental research designs in our context.

JEL Classification: C83, C93, J63, J64

Keywords: Hawthorne effect, panel conditioning, job search, labour market outcomes, field experiment

Corresponding author:
Gesine Stephan
Institute for Employment Research (IAB)
Regensburger Str. 104
90478 Nuremberg
Germany

E-mail: gesine.stephan@iab.de

* We are grateful for comments by Ruben Bach, Michael Cooper as well as participants of the yearly meetings of EALE (Bergen, 2024), BeWell (Magdeburg, 2024) and the Scottish Economic Society (Glasgow, 2024). The DIM unit of IAB, in particular Stephan Grießemer, provided crucial support in carrying out the sampling. We also appreciate financial support by the German Science Foundation (DFG) through grants EI 379/11-1, SCHO 1270/5-1, and STE 1424/4-1. The experiment underpinning the paper was approved by an ethics committee of Freie Universität Berlin as part of an overarching project (approval no. 169/2017) and pre-registered at the AEA registry (https://www.socialscienceregistry.org/trials/2657) in January 2018).

1. Introduction

In the 1920s and the 1930s, the US National Research Council conducted several experiments on workplace productivity at the Hawthorne plant of Western Electric (see, e.g., Levitt and List 2011). It is handed down that, to the surprise of the researchers, productivity changes were observed not only in the experimental group whose working conditions had been altered but also in the control group of workers whose conditions remained unchanged. This seemed to reveal that the mere awareness of being observed can lead to changes in behaviour, a phenomenon later termed 'Hawthorne effect'.[1] Such participant reactivity effects are a threat to the internal validity of study results: The information gathered is biased by the fact that the participants were surveyed. A related threat can arise in panel surveys, which are of paramount importance for investigating dynamic processes and estimating causal effects: Over time, repeated participation in surveys can induce changes in observable behaviour due to changes in either actual behaviour or in reporting biases, such as social desirability (Chadi 2013, Bach 2021, Cernat and Keusch 2022). If survey participation has an impact on such outcomes during later waves of a panel study, this is called panel conditioning. In general, panel-conditioning might occur from changes in behaviour or from changes in reporting (Bach 2021). By means of a field experiment, we study Hawthorne effects and changes-in-behaviour panel conditioning with respect to actual behaviour in job seekers who took part in a high-frequency survey.

Job search is an important process affecting a person's future income, job quality and work-life balance, among other things. Ending a period of insecure employment or unemployment is also key to improving health and well-being (Clark et al. 2008, Cygan-Rehm et al. 2017, Lawes et al. 2022, Reichert and Tauchmann 2017). This means researchers would have to consider profound ethical issues going beyond internal study validity if study participation was found to interfere with job search behaviour as a form of a Hawthorne effect. For instance, feeling closely monitored could make job seekers accept job offers 'too quickly', resulting in bad job quality. This might be particularly true for surveys run by institutions of the public sector, which also administer income support for the unemployed, such as public universities and governmental research departments. Moreover, job seekers may seek to align their behaviour with the social norm to work (e.g. Stutzer and Lalive 2004, Günther et al. 2024). Participating in a survey about job search, among other things, may remind respondents of their own norm non-compliance (Halpern-Manners et al. 2017). Overall, these are good reasons to expect survey participation to increase the probability of being employed in our context.

[1] Later research has revealed little evidence to suggest Hawthorne effects actually happened in the course of the Hawthorne experiments. Yet the anecdote and thus terminology prevail (Levitt and List 2011).

On the other hand, study participation as 'good-citizen' behaviour or contribution to the greater good could be seen as a way of compensating for a lack of job search effort and norm non-compliance (Groves et al. 1992, Misra et al. 2012). Where survey participation becomes rather time-consuming, it may reduce the time spent on job search, similar to the lock-in effect of program participation (e.g., Sianesi 2008). Both these issues would increase the individual's probability of unemployment and, hence, their dependency on public income support. This could work against the expected positive effect of survey participation on employment discussed above. In any event, it seems straightforward to assume that the intensity of study participation may amplify survey participation effects. In a longitudinal study, high participation intensity can originate from the frequency of measurements (e.g., monthly versus yearly) and the number of items that are to be answered at each measurement ('survey wave').

Studying the effect of survey participation on real-life outcomes involves at least two key challenges; finding an adequate control group; and measuring outcomes of interest independently from other changes in reported outcomes due to changes in reporting biases or panel attrition. When addressing the first challenge, the widely accepted gold standard is to randomly assign study participants to a control group surveyed only once or not at all (e.g. Axinn et al. 2015, Warren and Halpern-Manners 2012). This issue arises as the willingness to participate in surveys, as well as attrition over time in longitudinal studies, are non-random. Other approaches are to compare answers of longer-term panel participants with those of panel refreshers (Van Landeghem 2014), or the use of instrumental variables as quasi-random assignment of study participation (Bach and Eckman 2019).

In regards to panel studies, actual Hawthorne effects (i.e., changes in behaviour) need to be disentangled from changes in reporting behaviour, time trends, and other error sources such as interviewer effects (e.g., Das et al. 2011, Bach 2021). Thus, to resolve the second challenge, matching survey data with administrative records is considered the gold standard for identifying such outcomes, as administrative data are usually reported independently from the survey in question. Unfortunately, only a few studies have yet utilized combined survey and administrative data (e.g., Crossley et al. 2017, Bach and Eckman 2019). Alternatively, digital trace data may be used to investigate the impact of survey participation, however they come with substantial measurement challenges of their own (Bähr et al. 2020, Cernat and Keusch 2022).

We combine both gold standards to investigate if participation in a panel survey affects the labour market outcomes of the participants: First, we randomly assigned part of the individuals willing to participate in the German Job Search Panel (GJSP) to a *control group* that is excluded from the survey. In contrast, *treatment group* individuals continued to participate. The GJSP is a high-intensity panel survey following the same people for up to two years (Hetschko et al. 2022): Using an innovative survey app allowed for frequent measurements every month, including the experience sampling method, which required subjects to respond multiple times on one day each month (this is also known as momentary mood assessment; Stone and Litcher-Kelly 2006).[2] In addition, a structured diary technique was used as part of the GJSP every three months to document time use and emotions experienced over the course of the day (Kahneman et al. 2004). Additionally, hair cortisol levels were collected to gather a biomarker of chronic stress. This required individuals to send in samples of their hair every three months (Lawes et al. 2024a).

Second, we link the GJSP survey data with high-quality administrative data to compare the labour market results of the actual ('treated') survey participants to the not surveyed control group. The administrative data, the Integrated Employment Biographies (IEB, Frodermann et al. 2021), are provided by the German Federal Employment Agency (FEA, in German *Bundesagentur für Arbeit*) and contain comprehensive information about periods in employment, unemployment as well as participation in active labour market programs, among other things. We augment the data with additional information on whether job seekers attend meetings at their local branch of the FEA to measure their job search efforts.[3] None of these data can be influenced by attrition or changes in reporting behaviour in the GJSP.

Our contribution to the literature is at least threefold: First and foremost, we answer the question of whether participation in a high-frequency app-based survey had causal effects on the subsequent labour market outcomes of participating job seekers. Special emphasis is placed on employment transitions (e.g., ending a job, taking up a new job) and participation in active labour market policy programs (e.g., retraining). Much of the existing literature is related to the areas of voting, retirement savings, and health. It often confirms that participation indeed has a (context-dependent) impact on behaviour (for overviews see, e.g., Bach 2021, Cernat and Keusch 2022).[4] Three studies that also fulfil the two gold standards outlined above are Persson (2014), Crossley et al. (2017) and Zwane et al. (2011). Persson (2014) shows that being randomly assigned to participating in a high-intensity election survey prior to the date of the election increases turnout compared to participating after the election. Presumably this is because the survey triggered some interest in the election and/or increased the perceived social pressure to vote. Voting was measured by official register files. Crossley et al. (2017) implemented a random assignment to modules with detailed questions on needs in retirement within a population-representative internet panel.

[2] Recent work by Eisele et al. (2023) suggests there are reactivity effects of completing the experience sampling method, however not necessarily in the form of behavioral change. Previously, Lischetzke and Eid (2003) report that high attention to feeling can be beneficial to momentary well-being if individuals have strong mood regulation abilities, whereas it could be detrimental if mood regulation abilities are weak.
[3] These meetings take place between the job seeker and a staff member responsible for their case of job search. Job seekers are required to attend these meetings regularly to remain eligible for financial support.
[4] Moreover, we note there is a parallel literature in marketing research examining the effects of running consumer surveys on purchasing behaviour (e.g. Dong et al. 2014).

From administrative wealth data, they link information on actual savings. They find that households reacted to being confronted with retirement questions by reducing their non-housing saving rate. The authors' explanation for this finding is that surveyed individuals had a salience shock and realized that they indeed needed fewer savings. In a series of experiments, Zwane et al. (2011) find that being surveyed about health increased the demand for water treatment products and medical insurance, whereas being surveyed about borrowing behaviour did not influence the demand for a microloan.

We are aware of only one study focusing on labour market effects of panel participation, which did not conduct an experiment, however: Bach and Eckman (2019) use invitations to participate in the annual German Panel Labor Market and Social Security (PASS) as an instrument and find that participation in the panel led to increasing participation in active labour market programs. The authors acknowledge that their instrumental variable might not fully address all endogeneity concerns. Furthermore, they relate a specific population (welfare recipients) to a specific outcome (active labour market program participation). While our population is also specific (originally registered job seekers), our analyses cover a much broader range of labour market outcomes in our case, and survey participation was more invasive in that it occurred at a greater frequency (monthly instead of annually).

As a second contribution, we assess the need for an experimental design to identify the real-life effects of survey participation. One issue to consider here is that generating a control group of randomly excluded people who were actually willing to participate in the survey prolongs the recruitment phase and requires additional resources. We therefore compare the treatment group (willing to participate, not randomly excluded) to a further group of individuals who were invited to take part in the survey but did not respond to the survey invitation, the so-called *no-signup group*. We check if controlling for a vast range of observable characteristics allows us to identify the same effects of study participation we find in the experimental data.[5] This selection-on-observables approach is less costly in terms of time and other resources but comes at the risk of endogeneity bias in the estimated effect of survey participation on labour market outcomes due to unobserved characteristics.

We also note that existing research on reactivity effects often stems from surveys that are relatively non-invasive. This contrasts with the common belief that more demanding surveys, such as those requiring frequent, detailed measurements, are more likely to suffer from reactivity and other data quality issues (Gochmann et al. 2022, Eisele et al. 2023). As a third contribution to the survey-methodological research across disciplines, our analysis focuses on participation in a high-

[5] Note that we cannot replicate the approach of Bach and Eckman (2019) and use survey invitations as an instrument as we invited all registered job seekers supposed to be part of a mass layoff to participate in our survey (see also section 2).

intensity panel survey involving detailed monthly measurements (see above). By studying this type of survey, we contribute to a better understanding of reactivity effects in more demanding contexts.

Despite the highly intense participation in the GJSP, we do not find significant effects of panel survey participation on labour market outcomes. In our analysis, we correct for multiple testing, but we would also find no statistically and economically significant effects if we neglected this issue. This is reassuring in terms of internal study validity and ethical concerns. What is more, we find pronounced differences in transitions into training between survey participants and people not interested in participating, even after controlling for a vast number of observables and correcting for multiple testing. This points to the necessity of employing experimental designs for the analysis of reactivity effects.

In what follows, section 2 describes the experimental design and the data used in greater detail. Section 3 shows balancing results and outlines the methods used. Section 4 presents the empirical findings. Section 5 concludes.

2. Experimental design and data sources

Our field experiment took advantage of the collection of data from the German Job Search Panel (GJSP; see Hetschko et al. 2022 for a detailed data report).[6] The actual purpose of the dataset is to provide longitudinal survey data for examining the effects of job search and unemployment on well-being and health on a monthly basis. For this purpose, potential survey participants were drawn from not-yet unemployed job seekers who registered with the German FEA.[7] Among these cases, a sizable fraction of workers actually entered unemployment, whereas a similarly large share were able to stay in employment (Stephan 2016).[8] Persons who were identified as part of an upcoming mass layoff were oversampled, as it is usually acknowledged in the literature that losing a job during a mass layoff is uncorrelated with unobserved characteristics, unlike ending a job due to other reasons (e.g., Kassenboehmer and Haisken-De New 2009).[9] The sample was

[6] The first studies based on the GJSP data examined the effects of Covid-19 on the mental well-being of workers (Schmidtke et al. 2023) as well as the effects of unemployment on well-being (Lawes et al. 2023, 2024b) and hair cortisol (Lawes et al. 2022).
[7] In Germany, to avoid a cut-off period of unemployment benefits, individuals must register as job seekers with the Federal Employment Agency (FEA) three months in advance of the end of their employment relationship, or otherwise within three days of receiving their notice of dismissal.
[8] Oftentimes workers register with the FEA because their fixed-term contract expires. However, in many of these cases the contract is eventually extended or made permanent. Others expect their company to close down, or that they will be part of a mass layoff, which is then prevented at the last minute.
[9] Our definition of mass layoffs largely follows §17(1) of the German employment protection act (Kündigungsschutzgesetz): > 5 layoffs in plants with up to 59 employees, 10% in plants with 60-250 employees, > 25 layoffs in plants with 251-499 employees, ≥ 30 layoffs in plants with 500+ employees.

restricted to individuals with German citizenship in order to avoid language issues with the survey questionnaires.

From November 2017 to May 2019, persons of ages 18 to 59 years meeting the criteria described above were invited to take part in the online entry survey of the GJSP. This survey provided access to the survey app if a number of inclusion criteria were met. These criteria included a random group assignment for the purpose of our field experiment. Around two thirds randomly selected participants were invited to further participate in the survey. In the following, we refer to these persons as the *treatment group*. At a probability of one third, potential participants were excluded from further participation in the survey. These constitute our *control group*. Comparing the labour market outcomes of these two groups produces causal evidence about Hawthorne effects from GJSP participation.

As mentioned above, additional time and other resources were needed to conduct the experiment, mostly because the sample fills up more slowly if one excludes people willing to participate. It is thus worthwhile to test whether a selection-on-observables approach simply comparing the labour market outcomes of people unwilling to partake with those of the survey participants produces the same insights in regard to Hawthorne effects. We therefore compare a so-called *no-signup group* separately with the treatment group. These individuals were invited but did not participate in the entry survey. As is described in greater detail in Hetschko et al. (2022), their non-participation is non-random, and so we use this no-signup group in the analysis to find out about the scientific benefit of the costly field experiment.

Figure 1 provides an overview of the three groups and their roles in our study. After applying appropriate sample restrictions, our final sample comprises 1,524 persons in the treatment group, 803 persons in the control group, and 63,744 individuals in the no-signup group (see also Figure 1). In the Appendix, we document all sample restrictions in detail.

[Figure 1 around here]

For our analysis, we merge information from the GJSP entry survey and paradata on subsequent survey participation with data for all invited persons from the IEB (V16.00.01-202012; see Frodermann et al. 2021 for an IEB data report) and with data from the FEA meeting scheduling software (*Allgemeine Terminvereinbarung* ATV). The IEB contains administrative spell data (accurate to the day) on periods of employment subject to social security contributions, registered job search, unemployment or welfare benefit receipts, and participation in active labour market programs administered by the FEA.[10] ATV information provides us with data on scheduled, attended, and missed appointments of job seekers at their local FEA.

For data preparation of the IEB, we compute all individual and job characteristics on the day of signing up for the entry survey (which is known for the treatment group and the control group). Furthermore, we compute the previous and subsequent labour market history before and after the day of signing up. As the date is not available for the no-signup group, we compute a hypothetical signing up day for this group. To this end, we impute a hypothetical date of signing up for the no-signup group based on the mean number of days between job seeker registration and signup observed in the experimental sample (combined treatment and control group).

While a main focus of the survey was on different concepts of well-being and health, participants were also asked monthly about various socio-demographic characteristics, personality traits, coping resources, and their current labour market status.[11] If unemployed, they were asked, for instance, about their reemployment prospects, reservation wage, and job search activities. Employed individuals were asked about job characteristics, earnings, working hours and the likelihood of upcoming changes in their employment status. To spread out the burden of participation, different questionnaire modules would pop up on different days each month. Overall, we argue that taking part in the survey was a substantial burden on individuals in light of the high frequency of survey questionnaires to be completed, and the numerous questions to be answered each month.

Table A.1 in the Appendix shows the means of observed characteristics for the treatment and the control group, as well as the results from tests on equal means. To address the issue of multiple testing, we employ the Romano-Wolf multiple-hypothesis correction (Romano and Wolf 2005, 2016) using the Stata ado-file *rwolf* (Clarke et al. 2021), with 250 bootstrap replications performed. This correction method safeguards against the likelihood of erroneously rejecting one or more true null hypotheses within a group of hypotheses being examined in the same way. The procedure considers the actual dependence structure among the test statistics by means of resampling, leading to enhanced power in comparison to previous multiple-testing approaches such as the Bonferroni method. We consider basic socio-demographic characteristics, such as age, sex, and education, as well as the characteristics of the last job, belonging to the mass layoff sample, and the employment history over the last five years (e.g., years in employment subject to

[10] The outcome training is of particular interest due to the results by Bach and Eckman (2019). Around 90 percent of all active labor market programs taken up after (hypothetically) signing up are short trainings in a firm or at a private provider or longer lasting further training programmes.

[11] Measuring well-being alone made GJSP participation intense (for a complete account, see Hetschko et al. 2022). Monthly experience sampling (six measurements on one day) and quarterly day reconstructions were used to elicit momentary happiness and time use. Cognitive well-being and mental health data were also collected using multiple items. Several instruments measured eudaimonic well-being, including a 24-item version of the Ryff (1989) scales. On a quarterly basis, respondents were invited to send in samples of their hair for the measurement of the stress hormone cortisol (for details, see Lawes et al. 2024a).

social security contributions, in unemployment, and with benefit receipt). For categorical variables, none of the means differ between the treatment group and the control group at conventional levels of significance, confirming randomization success. Note that this also holds true if we do not correct for multiple testing. Table A.2 additionally displays results from Chi-Square tests for differences in the distribution of these variables, which are in line with the previous findings.

Table A.1 also shows the means of observed characteristics for the additional comparison group not signing up for the entry survey and results from multiple-hypothesis corrected tests on equal means between the treatment group and the no-signup group. Here, we do find significant differences for many characteristics. This may partly be due to the considerably larger sample and, thus, enhanced statistical power. However, the mean deviations from the treatment group are also larger for the no-signup group than for the control group. This confirms that participation in the GJSP was non-random (see, also, Hetschko et al. 2022). For instance, individuals signing up for the entry survey were, on average, better educated and worked in jobs with higher requirements than individuals in the no-signup group. This highlights that a simple mean comparison of the labour market results of the treatment group and the no-signup group would be confounded if interpreted as an effect of survey participation.

3. Estimating labour market outcomes

Hawthorne effects might take some time to arise or require repeated monthly participation. In particular due to panel attrition, however, potential effects may cease to be visible in the very long run. We present findings for six outcome variables. With respect to duration outcomes, van den Berg et al. (2024) discuss the challenges that randomized controlled trials encounter when dealing with survival outcomes and propose analysing unconditional probabilities of transitions within certain durations as the most appropriate method. As the authors outline, randomization is lost if the analysis is conditioned on survival at a specific time point, as the composition of survivors may vary within groups over time (see, also, Abbring and Van den Berg 2005). This implies that a competing risk analysis is unsuited for analysing data from a randomized controlled trial, as it requires censoring the data as soon as a transition into one competing state occurs.

We thus present results on three important unconditional labour market transitions and three outcomes that can be interpreted as job features or indicators of search effort, all measured until 360 days after (hypothetically) signing up for the survey.

1. We first investigate if individuals had a transition out of regular *employment* during the 360 days after (hypothetically) signing up.[12] In fact, this applies to half of all observations. Employment may take place in a continuing or new employment relationship. Indeed, many job seekers search successfully for a new job when expecting to terminate an employment relationship without ever entering unemployment. When computing the variables, we bridge gaps between two separate episodes of employment of up to 7 days to allow for short transition periods between jobs.
2. As a natural counterpart, we analyse if individuals entered *unemployment* after registration as job seeker. This variable is not an exact mirror of employment exits as a substantial share of individuals transition from employment into states other than unemployment.
3. Individuals who register as job seekers or are unemployed may take part in active labour market programs. We thus also examine transitions into *subsidized (short) training* during the 360 days after signing up.
4. As an indicator of employment quality, we compute *average daily earnings* within this period. For days without labour earnings, we impute a wage rate of zero.
5. As a job-related indicator of job search outcomes, we investigate if individuals took up a job in a *different municipality*.
6. As another aspect of search behaviour, we check if individuals had at least one *cancelled appointment* at their employment agency during the 360 days after (hypothetically) signing up.

Ideally, we would also have studied outcomes related to the GJSP's focus on well-being and health, but we naturally lack the corresponding data for the control group and the no-signup group.[13]

For each outcome, we estimate two specifications of linear probability models or OLS (for wages), respectively, to compare the treatment group separately with the control group and with the no-signup group. First, we include only a dummy variable in the estimates for the treatment group, which constitutes a simple comparison of means. Second, the OLS model controls for a

[12] We exclude periods of marginal employment, but include times of employment subject to social security contributions for which the FEA paid a wage subsidy. The latter does not have a sizeable impact on our results, as differences in both outcomes only appear in the third decimal place.

[13] Furthermore, not all pre-registered outcomes (duration of job search, relocation, commuting when reemployed, wage when reemployed, future unemployment probability, characteristics of future employer) could be examined. In particular, we decided not to investigate the duration of job search as registered job search might take place during times of employment as well as during times of unemployment and is therefore difficult to interpret. Instead, we added cancelled meetings with the employment agency as an alternative indicator of search effort. For mobility, we analyze changes in the address of the employer as information on the home address is partly not consistent between employer notifications and data from the operative systems of the FEA.

wide range of explaining variables. For a well-conducted field experiment, however, a comparison of means should already be sufficient to identify causal effects.

On the contrary, comparing the treatment group and the no-signup group, any estimated effects of survey participation might reflect the role of characteristics correlating with the willingness of singing up (e.g., education, see Hetschko et al. 2022). Hence, to identify actual effects of survey participation, all of these characteristics need to be observable and controlled for. While the former can only be assumed, the latter is carried out by means of the aforementioned OLS models. In addition to that, we present estimates using entropy balancing as a non-parametric way of controlling for observables (Hainmueller and Xu 2013). Here, observations in the no-signup group are reweighted upon the condition that they perfectly match the first and second moments of observables in the treatment group.

For further analyses of the treatment and control group, we include variables for the intensity of the treatment, measured by continued survey participation over at least 7 months, as well as participation in the additional cortisol study. In this context, we also discuss the possibility that attrition influences treatment effects via lowering treatment intensity and by being non-random. To uncover effect heterogeneity, we interact the survey dummy with belonging to the mass layoff sample, being female, having a temporary contract at the time of registering, and having had a recall to a previous employer during the last five years before signing up.[14]As we analyse several outcome variables, we again account for multiple testing by conducting the Romano-Wolf multiple-hypothesis correction with 250 bootstrap replications (Clarke et al. 2021). In the following, Tables 1 to 3 contain information on point estimates, uncorrected *p*-values (in parenthesis) as well as multiple-testing-corrected *p*-values (in braced brackets). We consider all estimates using the same specification and sample as a group of tested hypotheses. For instance, we consider the comparison of the treatment and control group across six outcomes in Table 1, Panel I, as one group of hypotheses as the sample and the specification are the same.

4. Empirical results

Labour market transitions

Table 1 presents our main set of results, displaying estimated coefficients for the survey variable from estimates with and without covariates (see Table A.1 in the Appendix). The reference group in the upper panel I of the Table is the control group. For this panel, estimated coefficients based on the full set of covariates are presented in Table A.3 in the Appendix. Panels II and III show treatment effects estimated when the no-signup group is used as the reference group, without and with entropy balancing. Additionally, Table 1 informs about the respective mean values of the

[14] A recall is defined as the start of an employment with a firm for which the employee worked before.

outcome variables for the reference groups. As our outcome variables have different means, estimated relative effects from the models controlling for the full set of covariates are additionally displayed in Figure 2. While statistical significance is shown in Table 1, Figure 2 thus provides additional information about the economic significance of estimated effects. To obtain the relative effects, coefficients from models with covariates in Panels I and II of Table 1 are divided by mean values for the control group.

[Table 1 and Figure 2 around here]

Our first set of outcome variables focuses on transitions until 360 after (hypothetically) signing up. In the control group, around 53 percent of all individuals had a transition out of employment, 44 percent entered unemployment, and about 13 percent participated in a (short) training program. Panel I of Table 1 shows that all three transitions do not differ significantly between the treatment and the control group. This holds true with and without controlling for covariates, with respect to both statistical and economic significance, and does not depend on correcting for multiple testing. Indeed, Figure 2 shows that the estimated relative effect sizes are generally of a small magnitude. This provides convincing evidence that participation in our demanding high-frequency survey did not have an impact on the labour market outcomes of participants.

Estimated effects for the no-signup group are displayed by constants in the middle panel II of Table 1. Recall that this non-randomised part of the study is used to assess the value of running the experiment as compared to a selection-on-observables approach (see also Figure 1 above). Including only the treatment variable but no further control variables, we find no significant differences in transitions (at least, once we correct for multiple-hypothesis-testing). Controlling for observable attributes of both groups, however, the results suggest a significantly positive effect of survey participation on transitions into training, even when correcting for multiple-hypothesis-testing. These estimates are also economically significant as they account for around 20 percent of the constant from models without covariates (as can be seen from Figure 2). These differences are also evident if we use entropy balancing to achieve similar distributions of observable characteristics in the no-signup group and the treatment group (Panel III of Table 1).

The most interesting result from this part of the analysis is that we find statistically and economically significantly more transitions in subsidized (short) training for the treatment group than for the no-signup group. This implies that at least some unobserved differences between the treatment and the no-signup group remain after controlling for observable characteristics, and that these unobserved differences are correlated with the propensity to participate in subsidized (short) training.

Job features and search effort

Even if we find no differences between labour market transitions, survey participation might still have an impact on job quality. Within 360 days after (hypothetical) random assignment, individuals in the control group realized on average daily wages of around 109 euros (imputing zeros for days without employment). Table 1 shows that earnings in the treatment group are around 4 euros per working day higher than in the control group (Panel I), but this difference is not significant. Differences are larger and strongly significant if we compare the treatment group with the no-signup group and include further covariates (Panel II). These changes, however, are no longer significant once we take the entire set of covariates into account and correct for multiple-hypothesis-testing or conduct entropy balancing (Panel III).

Most outcomes discussed above are not entirely controlled by the job seekers. For example, consider our finding that survey participation does not impact transitions out of regular employment. Here, survey participation might still increase the job search efforts of a newly registered job seeker, but not enough to actually improve job finding chances before entering unemployment, which also depends on labour market conditions. In the following, we therefore examine two more direct measures of job search efforts.

First, we investigate if a job seeker took up a job in a different municipality within 360 days of (hypothetically) signing up. This could be either seen as a job feature or an indicator of search effort, as individuals looked for a job in a broader regional context (compared to their previous job). On average, around one third of job seekers did indeed start working in a different municipality. However, we find no significant differences between the treatment and the control group. We do find differences between the treatment and the no-signup group (with and without controlling for covariates; see Panels II and III), but these become insignificant as soon as we correct for multiple testing.

Second, we check whether the individual job seeker had a scheduled appointment at their local FEA branch and whether that meeting took place. If a scheduled meeting did not take place, this is mostly due to the job seeker not showing up, which we interpret as a measure of lacking search effort. As the duration of both job search and of potential unemployment varies across individuals, we analyse if at least one appointment scheduled with the local employment agency did not take place within the 360 days after (hypothetically) signing up for the survey. In all three groups investigated, the share of individuals with at least one missed appointment was around one third. We find, however, no significant differences between the treatment and the control group as well as between the treatment and the no-signup group.

Treatment intensity and attrition

Another issue to consider in the context of our study is attrition. In principle, attrition might impact on treatment intensity. An extreme example would be a situation where all participants (i.e. the treated) drop out quickly after the random exclusion of the control group, implying a very weak treatment, clearly contradicting our claim of a high-intensity panel survey. While this is clearly not the case for the GJSP, the example shows that any attrition works against any treatment effect.[15]

To investigate this issue, we revisit our treatment effects for parts of the treatment group who were intensively treated. To this end, we interact survey participation with two indicators of treatment intensity. Of those entering the survey, 53 percent still participated in the seventh month (thus more than half a year). The vast majority of them completed 80 percent or more of the questionnaire (Hetschko et al. 2022). Furthermore, 26 percent participated in the cortisol study, which involved sending in strands of hair to receive an objective stress measure.[16] For the control group, both treatment intensity dummies were assigned a value of zero. Note that we hereby focus on the field experiment only as this is where we try to obtain causal evidence.

Table 2 presents estimates for the six outcome variables examined above, again controlling for the set of control variables described in Table A.1, and for 360 days after (hypothetically) signing up. The correction for multiple-hypothesis-testing is applied again, too, but this does not alter our conclusions. We find no statistically or economically significant effects for the intensity measures.

[Table 2 around here]

A related issue arises from the fact that attrition is a non-random process. Treatment group individuals who continued to participate are potentially different from those who dropped out. Hence, while the randomisation has ensured balanced samples at the point of signup, treated and control group observations potentially started to differ at any later point, obviously including month 7. Having said that, Hetschko et al. (2022) report little evidence for systematic differences between participants and non-participants even as late as month 7 across a variety of individual characteristics. Females seem more likely to stay on, however the effect is significant only at the 10% level. As we show further below, females do not differ, however, when it comes to the treatment effects. Overall, there is little evidence to suggest that our results are confounded by attrition.

[15] Given that even the control group completed a short part of the entry survey until exclusion one might argue that they were minimally treated, too. This makes the issue of attrition in the treatment group particularly relevant.

[16] Not every person that was willing to participate in the cortisol study could in fact participate. For example, participation required a minimum length of hair, which is why many male GJSP participants could not partake in the cortisol study. As a result of non-random nonresponse, the results based on this intensity indicator for treatment intensity should be interpreted cautiously.

[Table 2 around here]

Subgroup analysis

For a heterogeneity analysis of the experimental evidence (i.e. treatment vs control group), we interact the treatment group with registering due to a mass layoff, gender, having a temporary contract at the time of registering, and having experienced at least one recall during the last five years. These variables seem particularly interesting as the economic research on the effects of unemployment often restricts itself to mass layoffs, which are less prone to be correlated with individual unobserved characteristics in comparison to other types of job terminations (e.g., Schmieder et al. 2023). A gender-specific analysis seems appropriate as the labour market behaviour of men and women differs in many respects (e.g., Borella et al., 2023). Individuals on temporary contracts often have to register as job seekers due to institutional constraints, even if the chance of a contract extension is high (Stephan 2016). Furthermore, individuals who expect to be recalled have a smaller incentive to exert search effort.

The results are presented in Table 3, controlling for the full set of covariates (see Table 1) and correcting for multiple-hypothesis-testing. While most of the main effects do not have a significant impact on the outcome variables when we correct for multiple testing, individuals with a past recall left employment statistically and economically significantly earlier and less often took up a job in a different municipality. We find no significant interactions between gender, being on a temporary contract, or having had a recall, and survey participation. However, individuals taking part in the survey who were dismissed as part of a mass layoff seem to enter unemployment somewhat earlier than those who were dismissed for other reasons. For this group, survey participation appears to cancel out the fact that individuals experiencing a mass layoff are generally entering unemployment later than those who were dismissed for other reasons. Overall, however, we find little indication of heterogeneity across subgroups.

[Table 3 around here]

5. Conclusions

We investigate if participation in an innovative app-based survey on job search and well-being had an impact on labour market outcomes within a year of signing up for the survey. To this end, we combine two gold standards of survey research: First, we conduct a field experiment, where we randomly included one third of eligible individuals signing up for a survey from participation and use this group as a control group. Second, we merge information on survey participation with administrative information on labour market outcomes. This allows us to rule out that our results are in any way related to reporting errors.

Our most important finding is that participation in the survey, on average, had no impact on any of the investigated labour market transitions, namely out of employment, into unemployment, into subsidized (short) training, as well as daily wage rates, taking up a job in a different municipality and cancelled appointments with the local employment agency. These outcomes seem appropriate, given that our survey participants were initially registered job seekers and considering the potential influence of the survey on job search behaviour and subsequent labour market outcomes. Furthermore, we do not find an effect for different indicators of treatment intensity measured by duration of survey participation and involvement in an accompanying cortisol study. There is also little evidence for effect heterogeneity across subgroups, except for the finding that people who registered as job seekers amid a mass layoff enter unemployment earlier when participating in the panel survey.

In addition, we show that even controlling for a wide range of observable characteristics and correcting for multiple-hypothesis-testing, a comparison with individuals not signing up for the survey would have led to misleading conclusions. Regression results show that survey participants statistically and significantly more often take up subsidized (short) training if compared to the no signup group. Thus, there seems to be some remaining selection into survey participation based on unobservable characteristics, creating a false sense of an impact of survey participation in the regression. This reiterates the importance of experimental research designs for identifying effects in our context. In this sense, our field experiment was worth the effort, even though excluding the control group from the survey meant we had to spend more time and resources to fill up our sample.

Notwithstanding the caveat of a selective population under consideration, our findings provide good news for survey researchers especially in the area of labour economics. The lack of reactivity effects speaks to the internal validity of research results obtained from analysing survey data. This is in spite of the fact that the examined data collection is highly demanding when it comes to the frequency and scope of the repeated measurements. Comparing our findings to the previous literature (see section 1) indicates that the occurrence of Hawthorne effects depends on the particular circumstances of the survey in question, such as the area of study and the outcomes and subgroups analysed. There is also a possibility that Hawthorne effects are overstated where the analysis is not based on a causal research design and no correction for multiple testing is applied. Further research in this area should be conducted to get a more complete picture of when reactivity (does not) occur(s). This applies especially to the field of labour market research, where empirical studies of potential Hawthorne effects are still scarce.

References

Abbring, J. H., Van den Berg, G. J. (2005). Social experiments and instrumental variables with duration outcomes, Tinbergen Institute Discussion Paper TI 05-047/3.

Axinn, W. G., Jennings, E. A., Couper, M. P. (2015). Response of sensitive behaviors to frequent measurement. Social Science Research, 49, 1-15.

Bach, R. (2021). A methodological framework for the analysis of panel conditioning effects. In Cernat, A., Sakshaug, J. W. (eds), Measurement Error in Longitudinal Data (pp. 19–42). Oxford: Oxford University Press.

Bach, R. L., Eckman, S. (2019). Participating in a panel survey changes respondents' labour market behaviour. Journal of the Royal Statistical Society. Series A, Statistics in Society, 182(1), 263-281.

Bähr, S., Haas, G. C., Keusch, F., Kreuter, F., & Trappmann, M. (2022). Missing data and other measurement quality issues in mobile geolocation sensor data. *Social Science Computer Review*, 40(1), 212-235.

Borella, M., De Nardi, M., & Yang, F. (2023). Are marriage-related taxes and social security benefits holding back female labour supply?. Review of Economic Studies, 90(1), 102-131.

Cernat, A., & Keusch, F. (2022). Do surveys change behaviour? Insights from digital trace data. International Journal of Social Research Methodology, 25(1), 79–90

Chadi, A. (2013). The role of interviewer encounters in panel responses on life satisfaction. Economics Letters, 121(3), 550-554.

Clark, A., Diener, E., Georgellis, Y. & Lucas, R. (2008). Lags and leads in life satisfaction: A test of the baseline hypothesis. Economic Journal, 118(529), F222-F243.

Clarke, D., Romano, J., & Wolf, M. (2021). The Romano-Wolf multiple-testing correction in Stata. The Stata journal, 20(4).

Crossley, T. F., de Bresser, J., Delaney, L., & Winter, J. (2017). Can survey participation alter household saving behaviour? The Economic Journal, 127(606), 2332-2357.

Cygan-Rehm, K., Kuehnle, D. & Oberfichtner, M. (2017). Bounding the causal effect of unemployment on mental health: Nonparametric evidence from four countries. Health Economics 26(12), 1844-1861.

Das, M., Toepoel, V., & Van Soest, A. (2011). Nonparametric tests of panel conditioning and attrition bias in panel surveys. Sociological Methods & Research, 40(1), 32-56.

Dong, X., Janakiraman, R., & Xie, Y. (2014). The effect of survey participation on consumer behavior: the moderating role of marketing communication. Marketing Science, 33(4), 567-585.

Eisele, G., Vachon, H., Lafit, G., Tuyaerts, D., Houben, M., Kuppens, P., Myin-Germeys, I., & Viechtbauer, W. (2023). A mixed-method investigation into measurement reactivity to the experience sampling method: The role of sampling protocol and individual characteristics. Psychological Assessment, 35(1), 68-81.

Frodermann, C., Schmucker, A., Seth, S., vom Berge, P. (2021). Sample of Integrated Labour Market Biographies (SIAB) 1975 – 2019. FDZ Datenreport 01/2021.

Gochmann, V., Ohly, S., & Kotte, S. (2022). Diary studies, a double-edged sword? An experimental exploration of possible distortions due to daily reporting of social interactions. *Journal of Organizational Behavior*, *43*(7), 1209-1223.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. Public opinion quarterly, 56(4), 475-495.

Günther, T., Conradi, J., & Hetschko, C. (2024). Socialism, identity and the well-being of unemployed women. CESifo Working Paper 11154.

Hainmueller J., Xu Y. (2013). Ebalance: A Stata package for entropy balancing. Journal of Statistical Software, 54(7).

Halpern-Manners, A., Warren, J. R., & Torche, F. (2017). Panel conditioning in the general social survey. Sociological Methods & Research, 46(1), 103-124.

Hetschko, C., Schmidtke, J., Eid, M., Lawes, M., Schöb, R., & Stephan, G. (2022). The German Job Search Panel. OSF Preprints. https://doi.org/10.31219/osf.io/7jazr.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method, Science, 306(5702).

Kassenboehmer, S. & Haisken-DeNew, J. (2009). You're fired! The causal negative effect of entry unemployment on life satisfaction. Economic Journal, 119(536), 448-462.

Lawes, M., Hetschko, C., Schöb, R., Stephan, G., & Eid, M. (2022). Unemployment and hair cortisol as a biomarker of chronic stress. Scientific Reports, 12(1), 21573.

Lawes, M., Hetschko, C., Schöb, R., Stephan, G., & Eid, M. (2023). The impact of unemployment on cognitive, affective, and eudaimonic well-being facets: Investigating immediate effects and short-term adaptation. Journal of personality and social psychology, 124(3), 659.

Lawes, M., Hetschko, C., Sakshaug, J. W., & Eid, M. (2024a). Collecting hair samples in online panel surveys: Participation rates, selective participation, and effects on attrition. Survey Research Methods (online first).

Lawes, M., Hetschko, C., Schöb, R., Stephan, G., & Eid, M. (2024b). Examining interindividual differences in unemployment-related changes in subjective well-being: The role of psychological well-being and re-employment expectations. European Journal of Personality (online first).

Levitt, S. D., List, J. A. (2011). Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. American Economic Journal: Applied Economics, 3(1), 224-238.

Lischetzke, T., & Eid, M. (2003). Is attention to feelings beneficial or detrimental to affective well-being? Mood regulation as a moderator variable. Emotion, 3(4), 361-377.

Misra, S., Stokols, D., & Marino, A. H. (2012). Using norm–based appeals to increase response rates in evaluation research: A field experiment. American Journal of Evaluation, 33(1), 88-98.

Persson, M. (2014). Does survey participation increase voter turnout? Re-examining the Hawthorne effect in the Swedish National Election Studies. Political Science Research and Methods, 2(2), 297-307.

Reichert, A. & Tauchmann, H. (2017). Workforce reduction, subjective job insecurity, and mental health, Journal of Economic Behavior & Organization, 133, 187-212.

Romano, J. P., & Wolf, M. (2005). Exact and approximate stepdown methods for multiple-hypothesis-testing. Journal of the American Statistical Association, 100(469), 94-108.

Romano, J. P., & Wolf, M. (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. Statistics & Probability Letters, 113, 38-40.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. Journal of Personality and Social Psychology, 57(6), 1069-1081.

Schmidtke, J., Hetschko, C., Schöb, R., Stephan, G., Eid, M., & Lawes, M. (2023). Does worker well-being adapt to a pandemic? An event study based on high-frequency panel data. Review of Income and Wealth (online first).

Schmieder, J. F., Von Wachter, T., Heining, J. (2023). The costs of job displacement over the business cycle and its sources: evidence from Germany. American Economic Review, 113(5), 1208-1254.

Sianesi, B. (2008). Differential effects of active labour market programs for the unemployed. Labour Economics, 15(3), 370-399.

Stephan, G. (2016). Arbeitsuchend, aber (noch) nicht arbeitslos: Was kommt nach der Meldung? WSI-Mitteilungen, 69(4), 292-299.

Stone, A. A., & Litcher-Kelly, L. (2006). Momentary capture of real-world data. In M. Eid & E. Diener (Eds.), Handbook of multimethod measurement in psychology (pp. 61–72). American Psychological Association.

Stutzer, A., & Lalive, R. (2004). The role of social work norms in job searching and subjective well-being. Journal of the European Economic Association, 2(4), 696-719.

Van den Berg, G. J., Hofmann, B., Stephan, G., Uhlendorff, A. (2024, forthcoming). Mandatory integration agreements for unemployed job-seekers: A randomized controlled field experiment in Germany. International Economic Review.

Van Landeghem, B. (2014). A test based on panel refreshments for panel conditioning in stated utility measures. Economics Letters, 124(2), 236-238.

Warren, J. R., & Halpern-Manners, A. (2012). Panel conditioning in longitudinal social science surveys. Sociological Methods & Research, 41(4), 491-534.

Zwane, A.P., Zinman, J., Van Dusen, E., Pariente, W.; Null, C., Miguel, E., Kremer, M., Karlan, D.S., Hornbeck, R., Giné, X., Duflo, E., Devoto, F., Crepon, B. & Banerjee, A. (2011). Being surveyed can change later behavior and related parameter estimates, Proceedings of the National Academy of Sciences 108(5), 1821-1826.

Figures

Figure 1 Overview of the studied samples and timeline of the study

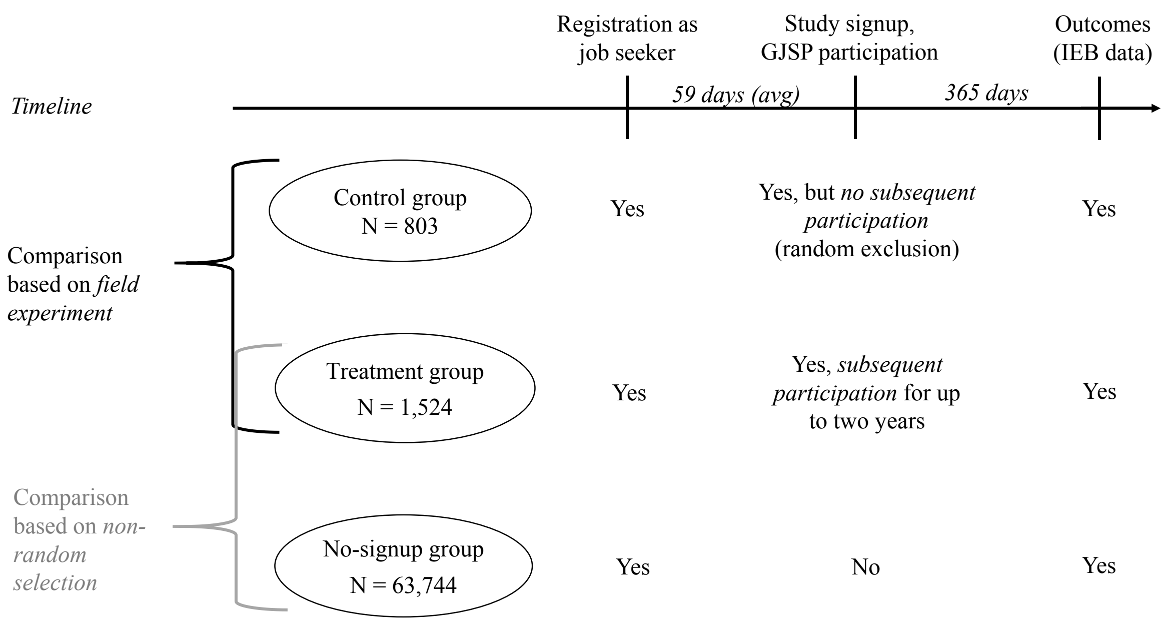
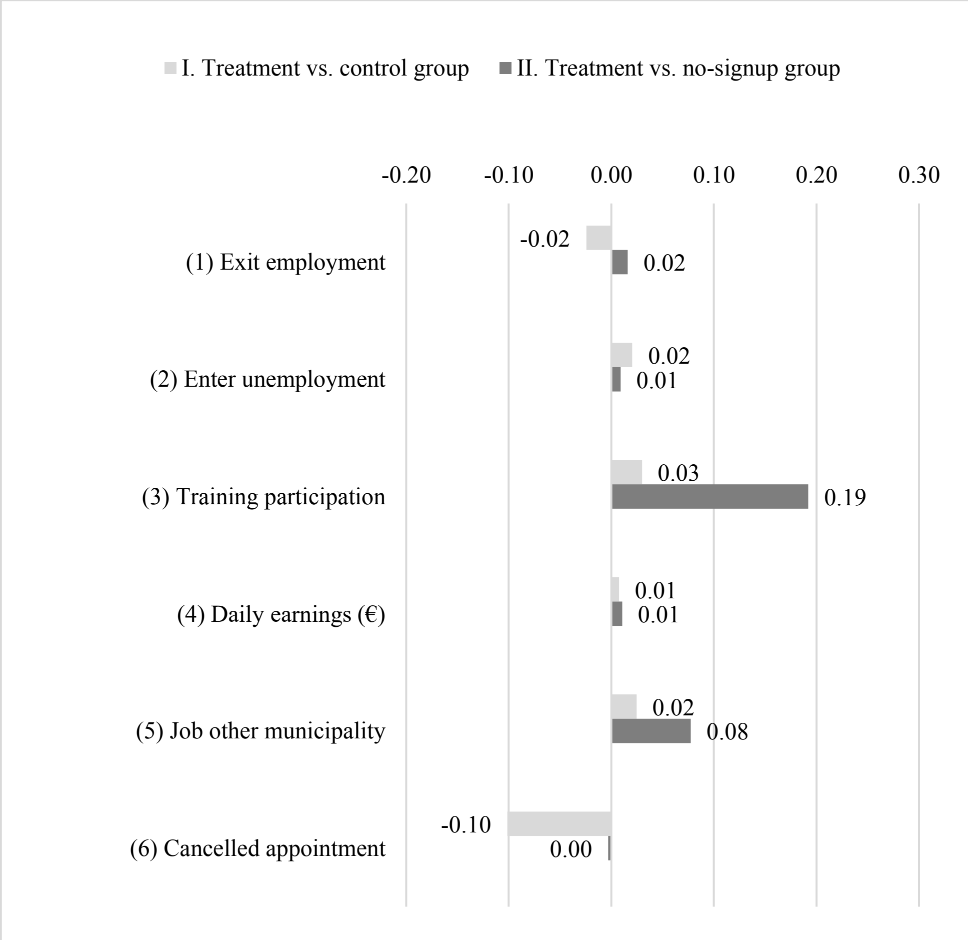


Figure 2 Estimated relative effects of survey participation compared to being in the control group or in the no-signup group, 360 days after (hypothetically) signing up



Source: GJSP and IEB (V16.00.01-202012).
Notes: Estimated coefficients from model with covariates divided through constant from model without covariates (Table 1).

Tables

Table 1 Estimated effects of survey participation compared to being in the control group or in the no-signup group on labour market outcomes until 360 days after (hypothetically) signing up

Coefficients, uncorrected *p*-values (in parentheses) and multiple-hypothesis corrected *p*-values (in braced brackets)

	Transitions			Job features and search indicators		
	(1) Exit employment	(2) Enter unemployment	(3) Training participation	(4) Daily earnings (€)	(5) Job other municipality	(6) Cancelled appointment
I. Treatment vs. control group						
Treatment group (1=yes)	-0.013	0.012	0.008	3.607	0.016	-0.034
	(0.54)	(0.59)	(0.60)	(0.08)	(0.43)	(0.09)
	{0.90}	{0.90}	{0.90}	{0.38}	{0.90}	{0.38}
Control variables	no	no	no	no	no	no
Treatment group (1=yes)	-0.013	0.009	0.004	0.812	0.008	-0.034
	(0.55)	(0.68)	(0.81)	(0.33)	(0.69)	(0.09)
	{0.95}	{0.96}	{0.96}	{0.82}	{0.96}	{0.46}
Control variables	yes	yes	yes	yes	yes	yes
Mean control group	0.533	0.440	0.134	108.606	0.325	0.335
II. Treatment vs. no-signup group						
Treatment group (1=yes)	0.018	0.007	0.018	15.702	0.032	-0.015
	(0.17)	(0.57)	(0.04)	(0.00)	(0.01)	(0.23)
	{0.37}	{0.59}	{0.12}	{0.00}	{0.02}	{0.41}
Control variables	no	no	no	no	no	no
Treatment group (1=yes)	0.008	0.004	0.024	1.013	0.024	-0.001
	(0.53)	(0.76)	(0.00)	(0.02)	(0.04)	(0.94)
	{0.88}	{0.96}	{0.05}	{0.16}	{0.19}	{0.96}
Control variables	yes	yes	yes	yes	yes	yes
Mean no-signup group	0.502	0.444	0.125	96.512	0.31	0.315
III. Treatment vs. no-signup group, with entropy balancing						
Treatment group (1=yes)	0.007	0.003	0.025	0.996	0.026	0.001
	(0.59)	(0.84)	(0.01)	(0.43)	(0.03)	(0.90)
	{0.91}	{0.98}	{0.05}	{0.88}	{0.14}	{0.98}
Control variables	no	no	no	no	no	no
Treatment group (1=yes)	0.007	0.003	0.025	0.933	0.026	0.002
	(0.58)	(0.83)	(0.01)	(0.06)	(0.03)	(0.89)
	{0.86}	{0.97}	{0.04}	{0.21}	{0.15}	{0.97}
Control variables	yes	yes	yes	yes	yes	yes
Mean no-signup group	0.513	0.449	0.118	111.218	0.315	0.299

Source: GJSP and IEB (V16.00.01-202012.

Notes: Linear probability models / OLS. List of control variables: See Table A.1. Corrected *p*-values are computed separately for Panels I to III, using the Romano-Wolf correction for multiple-hypothesis-testing (Clarke et al. 2021).

Table 2 Estimated effects of survey participation intensity compared to being in the control group on labour market outcomes until 360 days after (hypothetically) signing up

Coefficients, uncorrected p-values (in parentheses) and multiple-hypothesis corrected p-values (in braced brackets)

	Transitions			Job features and search indicators		
Treatment vs. control group	(1) Exit employment	(2) Enter unemployment	(3) Training participation	(4) Daily earnings (€)	(5) Job other municipality	(6) Cancelled appointment
I. Continued survey participation						
Treatment group (1=yes)	-0.011	0.023	0.005	0.964	-0.001	-0.021
	(0.66)	(0.36)	(0.79)	(0.32)	(0.95)	(0.38)
	{0.95}	{0.86}	{0.95}	{0.85}	{0.98}	{0.86}
Participated at least until month 7	-0.004	-0.027	-0.002	-0.285	0.018	-0.024
(1=yes)	(0.89)	(0.29)	(0.90)	(0.77)	(0.46)	(0.32)
	{0.98}	{0.87}	{0.98}	{0.98}	{0.93}	{0.88}
Control variables	yes	yes	yes	yes	yes	yes
II. Cortisol study participation						
Treatment group (1=yes)	-0.016	0.009	0.003	0.8	-0.004	-0.033
	(0.48)	(0.70)	(0.86)	(0.36)	(0.86)	(0.12)
	{0.88}	{0.97}	{0.97}	{0.85}	{0.97}	{0.48}
Participated in cortisol study	0.012	0.001	0.003	0.044	0.046	-0.001
(1=yes)	(0.69)	(0.96)	(0.89)	(0.97)	(0.09)	(0.96)
	{0.99}	{1.00}	{1.00}	{1.00}	{0.47}	{1.00}
Control variables	yes	yes	yes	yes	yes	yes

Source: GJSP and IEB (V16.00.01-202012). Notes: Linear probability models / OLS. List of control variables: See Table A.1. Corrected p-values are computed separately for Panels I and II, using the Romano-Wolf correction for multiple-hypothesis-testing (Clarke et al. 2021).

Table 3 Heterogeneous effects of survey participation intensity compared to being in the control group on labour market outcomes until 360 days after (hypothetically) signing up

Coefficients, uncorrected estimated p-values (in parentheses) and multiple-hypothesis corrected p-values (in braced brackets)

	Transitions			**Job features and search indicators**		
Treatment vs. control group	(1) Exit employment	(2) Enter unemployment	(3) Training participation	(4) Daily earnings (€)	(5) Job other municipality	(6) Cancelled appointment
Treatment group (1=yes)	0.042	0.041	0.018	3.949	0.015	-0.029
	(0.38)	(0.40)	(0.60)	(0.04)	(0.75)	(0.52)
	{0.88}	{0.88}	{0.90}	{0.17}	{0.90}	{0.90}
Treatment group	-0.097	-0.116	-0.014	0.088	-0.037	-0.013
* mass layoff	(0.03)	(0.01)	(0.65)	(0.96)	(0.38)	(0.77)
	{0.15}	{0.05}	{0.94}	{0.98}	{0.76}	{0.95}
Treatment group	0.045	0.081	-0.029	-2.395	0.022	0.039
* female	(0.30)	(0.06)	(0.35)	(0.15)	(0.60)	(0.35)
	{0.71}	{0.29}	{0.71}	{0.54}	{0.71}	{0.71}
Treatment group	-0.019	0.009	0.006	-1.904	0.003	-0.034
* temporary contract	(0.67)	(0.84)	(0.86)	(0.27)	(0.95)	(0.43)
	{0.97}	{1.00}	{1.00}	{0.83}	{1.00}	{0.89}
Treatment group	-0.03	-0.036	0.031	-3.051	0.014	0.014
* recall during last 5 years	(0.56)	(0.49)	(0.40)	(0.13)	(0.77)	(0.78)
	{0.91}	{0.91}	{0.89}	{0.46}	{0.95}	{0.95}
Mass layoff (1=yes)	0.064	0.082	0.004	0.300	0.005	0.012
	(0.09)	(0.03)	(0.89)	(0.84)	(0.90)	(0.75)
	{0.35}	{0.13}	{1.00}	{1.00}	{1.00}	{1.00}
Female (1=yes)	-0.054	-0.089	0.012	-0.821	-0.067	-0.059
	(0.14)	(0.02)	(0.63)	(0.56)	(0.05)	(0.09)
	{0.33}	{0.06}	{0.78}	{0.78}	{0.20}	{0.29}
Temporary contract (1=yes)	-0.043	-0.026	-0.054	1.901	-0.071	0.059
	(0.28)	(0.51)	(0.06)	(0.21)	(0.06)	(0.12)
	{0.43}	{0.53}	{0.17}	{0.43}	{0.16}	{0.27}
Recall during last 5 years (1=yes)	0.172	0.052	-0.051	-1.058	-0.121	-0.022
	(0.00)	(0.21)	(0.08)	(0.51)	(0.00)	(0.57)
	{0.00}	{0.51}	{0.24}	{0.76}	{0.01}	{0.76}
Control variables	yes	yes	yes	yes	yes	yes

Source: GJSP and IEB (V16.00.01-202012).

Notes: Linear probability models / OLS. List of further control variables: See Table A.1. Corrected p-values are computed using the Romano-Wolf correction for multiple-hypothesis-testing (Clarke et al. 2021)

Appendix

Sample restrictions

Out of 127,201 persons who were invited to take part in the online entry survey of the GJSP, 4,698 persons signed up for the entry survey (see Hetschko et al. 2022, for details).[17] Of those starting to participate in the entry survey, 2,747 persons fulfilled all substantive criteria (i.e., other than the random assignment) for further participation in the survey and used the app at least once.[18] 940 randomly chosen subjects of the 2,747 workers who signed up were excluded for the purpose of our field experiment. The remaining 1,873 randomly selected participants were invited to further participate in the survey. 122,503 persons who were invited did not sign up for the entry survey.

Based on the IEB information, we include only the focus group of the GJSP in our analysis sample, namely German individuals who were regularly employed at the date of signing up and at least half a year of tenure at their current employer. This excludes disproportionally many individuals from the no-signup group, as they entered unemployment or started a new job between being invited to participate in the GJSP and the hypothetical signup date. One reason might be that our invitation letter made clear our sole interest in 'still-employed' job seekers. This reiterates the non-random nature of the no-signup group in contrast to the control group when compared to the treatment group.

Individuals younger than 20 and older than 59 years at the date of (hypothetically) signing up for the survey were also not considered as the control group lacks any 18 or 19-year-old job seekers. For data preparation, we exclude employment spells with unrealistically low wages below 5 euros per day and impute missing values of the education variable based on entries in previous spells of a person. A small number of individuals are excluded as they could not be found in the IEB or information on their education is missing even after the imputation procedure.

Our final analysis sample then consists of 1,524 persons in the treatment group, 803 persons in the control group, and 63,744 individuals in the no-signup group (see also Figure 1).

[17] The sample used here is identical to what is described in Hetschko et al. (2022). Figures might slightly differ from other analyses based on the GJSP due to specific strategies of dealing with a small number of people who were invited more than once or who potentially falsely claimed to be still employed at signup.
[18] We exclude all individuals that did not submit the entry survey (246), were already unemployed (1,424) or on job probation (15), never used the app (35), or mistakenly took part in the survey (31).

Additional tables and figures

Table A.1 Means of observed characteristics and *p*-values from Romano-Wolf-corrected tests on equal means

	Treatment group (S)	Control group (C)	No signup group (N)	Corrected *p*-value S and C	S and N
Mass layoff sample (1=yes)	0.62	0.63	0.65	1.00	0.20
Gender (1=female)	0.52	0.55	0.44	0.97	0.00
East Germany (1=yes)	0.19	0.22	0.20	0.97	0.41
Education					
No occupational degree (1=yes)	0.02	0.03	0.08	0.97	0.00
Occupational degree (1=yes)	0.50	0.52	0.72	1.00	0.00
University degree (1=yes)	0.48	0.46	0.20	1.00	0.00
Characteristics last job					
Daily wage (in euros)	109	107	96	1.00	0.00
Temporary contract (1=yes)	0.60	0.63	0.53	0.92	0.00
Part-time (1=yes)	0.35	0.36	0.27	1.00	0.00
Age group					
20-29 years old (1=yes)	0.20	0.23	0.19	0.92	0.39
30-39 years old (1=yes)	0.38	0.38	0.28	1.00	0.00
40-49 years old (1=yes)	0.22	0.18	0.23	0.65	0.39
50-59 years old (1=yes)	0.20	0.21	0.30	1.00	0.00
Sector last job					
Manufacturing (1=yes)	0.15	0.13	0.25	1.00	0.00
Trade, maintenance, repair (1=yes)	0.09	0.10	0.12	1.00	0.01
Transport and storage (1=yes)	0.03	0.04	0.06	1.00	0.00
Information and communication (1=yes)	0.05	0.03	0.02	0.92	0.00
Scientific and technical services (1=yes)	0.08	0.10	0.05	1.00	0.00
Other business services (1=yes)	0.04	0.05	0.06	0.97	0.04
Public administration, defence (1=yes)	0.04	0.05	0.03	1.00	0.04
Education (1=yes)	0.24	0.22	0.10	1.00	0.00
Health and social care (1=yes)	0.14	0.15	0.10	1.00	0.00
Other services, private households (1=yes)	0.03	0.03	0.02	1.00	0.12
Temporary agency work (1=yes)	0.04	0.05	0.07	1.00	0.00
Other sector (1=yes)	0.06	0.06	0.11	1.00	0.00
Position last job					
Helper job (1=yes)	0.11	0.13	0.24	1.00	0.00
Professional job (1=yes)	0.36	0.40	0.51	0.96	0.00
Complex specialist job (1=yes)	0.15	0.12	0.09	0.79	0.00
Highly complex job (1=yes)	0.38	0.36	0.16	1.00	0.00
Employment history last 5 years					
Regular employment (in years)	3.93	3.85	4.01	0.97	0.12
With last employer (in years)	2.78	2.75	2.70	1.00	0.20
Unemployment (in years)	0.21	0.22	0.28	1.00	0.00
Unemployment benefits (in years)	0.18	0.15	0.21	0.97	0.01
Welfare benefit receipt (in years)	0.17	0.20	0.24	1.00	0.01
Recall (1 = yes)	0.21	0.24	0.25	0.89	0.02
Active labour market program (1 = yes)	0.16	0.16	0.19	1.00	0.04

Table A.1 continued

	Treatment group (S)	Control group (C)	No signup group (N)	Corrected p-value S and C	S and N
Signup quartile					
1st quartile 2018 (1=yes)	0.04	0.05	0.09	1.00	0.00
2nd quartile 2018 (1=yes)	0.09	0.10	0.10	0.97	0.20
3rd quartile 2018 (1=yes)	0.15	0.13	0.14	1.00	0.39
4th quartile 2018 (1=yes)	0.29	0.28	0.32	1.00	0.09
1st quartile 2019 (1=yes)	0.22	0.24	0.17	1.00	0.01
2nd quartile 2019 or later (1=yes)	0.22	0.20	0.18	1.00	0.01
Number of observations	1,524	803	63,740		

Source: GJSP and IEB (V16.00.01-202012)
Note: p-values are computed using the Romano-Wolf correction for multiple-hypothesis-testing (Clarke et al. 2021).

Table A.2 *p*-values from Chi Square tests for differences in the distributions of categorical variables

	p-values S and C	S and N
Education	0.19	0.00
Age groups	0.10	0.00
Sector last job	0.58	0.00
Position last job	0.06	0.00
Signup quartile	0.27	0.00

Source: GJSP and IEB (V16.00.01-202012)

Table A.3 Full regression results for outcomes until 360 days since (hypothetically) signing up for the treatment and the control group
Coefficients and uncorrected *p*-values (in parentheses)

	Transitions			Job features and search indicators		
Treatment vs. control group	(1) Exit employment	(2) Enter unemployment	(3) Training participation	(4) Daily earnings (€)	(5) Job other municipality	(6) Cancelled appointment
Treatment group	-0.013	0.009	0.004	0.812	0.008	-0.034
(1 = yes)	(0.55)	(0.68)	(0.81)	(0.33)	(0.69)	(0.09)
Mass layoff sample (1=yes)	0.001	0.007	-0.006	0.333	-0.02	0.004
(1 = yes)	(0.97)	(0.78)	(0.71)	(0.72)	(0.38)	(0.88)
Female	-0.025	-0.037	-0.007	-2.386	-0.054	-0.034
(1 = yes)	(0.25)	(0.10)	(0.66)	(0.01)	(0.01)	(0.11)
East Germany	-0.053	-0.040	-0.015	-2.394	-0.043	0.035
(1 = yes)	(0.04)	(0.13)	(0.43)	(0.02)	(0.08)	(0.15)
Education (reference: occupational degree)						
No occupational degree	0.076	0.087	0.061	-2.788	-0.137	0.043
(1 = yes)	(0.29)	(0.23)	(0.23)	(0.31)	(0.04)	(0.53)
University degree	0.101	0.102	0.031	6.166	0.019	-0.035
(1 = yes)	(0.00)	(0.00)	(0.12)	(0.00)	(0.48)	(0.19)
Characteristics last job						
Daily wage during last job	0.000	0.000	0.000	0.829	0.000	0.000
(in euros)	(0.24)	(0.17)	(0.42)	(0.00)	(0.64)	(0.13)
Temporary contract	-0.060	-0.025	-0.050	0.658	-0.071	0.035
(1 = yes)	(0.02)	(0.35)	(0.01)	(0.51)	(0.00)	(0.15)
Part time	0.088	0.101	0.044	-1.466	0.014	-0.006
(1 = yes)	(0.00)	(0.00)	(0.02)	(0.16)	(0.59)	(0.80)
Age group (reference: up to 30)						
30-39 years old	-0.018	-0.016	0.007	-1.527	-0.049	0.029
(1 = yes)	(0.56)	(0.60)	(0.76)	(0.20)	(0.10)	(0.31)
40-49 years old	-0.083	-0.053	0.013	-3.219	-0.073	0.019
(1 = yes)	(0.02)	(0.14)	(0.61)	(0.02)	(0.03)	(0.58)
50-59 years old	-0.018	-0.042	0.012	-4.108	-0.125	0.06
(1 = yes)	(0.62)	(0.25)	(0.63)	(0.00)	(0.00)	(0.08)
Employment history last 5 years						
Regular employment	-0.027	-0.022	-0.004	1.166	0.012	-0.034
(in years)	(0.02)	(0.05)	(0.60)	(0.01)	(0.26)	(0.00)
With last employer	0.028	0.038	0.020	0.424	0.018	0.023
(in years)	(0.00)	(0.00)	(0.00)	(0.22)	(0.03)	(0.01)
Unemployment	0.064	0.101	0.041	-0.217	0.057	0.015
(in years)	(0.10)	(0.01)	(0.14)	(0.89)	(0.12)	(0.68)
Unemployment benefits	0.007	0.007	-0.021	1.791	-0.060	0.01
(in years)	(0.88)	(0.88)	(0.51)	(0.31)	(0.17)	(0.82)
Welfare benefit receipt	-0.011	-0.003	-0.004	0.023	-0.034	0.015
(in years)	(0.62)	(0.91)	(0.78)	(0.98)	(0.10)	(0.46)
Recall (1 = yes)	0.153	0.029	-0.031	-2.932	-0.112	-0.014
(1 = yes)	(0.00)	(0.26)	(0.09)	(0.00)	(0.00)	(0.56)
Active labour market program	0.006	-0.011	0.033	-1.623	-0.001	0.05
(1 = yes)	(0.86)	(0.75)	(0.17)	(0.21)	(0.97)	(0.12)

Table A.2 continued

Treatment vs. control group	Transitions: *Enter un-em-ployment*	*Exit job*	Enter training	Job features and search indicators: Daily earnings (€)	Cancelled appointment	Job in other municipality
Sector last job (reference: manufacturing)						
Trade, maintenance, repair	0.027	0.017	0.025	-3.792	0.025	0.053
(1 = yes)	(0.54)	(0.70)	(0.41)	(0.03)	(0.55)	(0.20)
Transport and storage	-0.054	-0.111	-0.105	-2.031	-0.018	-0.034
(1 = yes)	(0.37)	(0.07)	(0.01)	(0.38)	(0.75)	(0.55)
Information & communication	0.076	0.121	-0.062	7.843	-0.029	0.047
(1=yes)	(0.19)	(0.04)	(0.13)	(0.00)	(0.59)	(0.39)
Scientific & technical services	-0.083	-0.070	-0.132	1.430	0.012	0.015
(1=yes)	(0.07)	(0.12)	(0.00)	(0.41)	(0.77)	(0.73)
Other business services	-0.004	-0.042	-0.022	-2.474	0.048	-0.041
(1 = yes)	(0.94)	(0.46)	(0.57)	(0.25)	(0.37)	(0.44)
Public administration, defence	-0.066	-0.106	-0.145	-1.547	-0.198	-0.027
(1=yes)	(0.25)	(0.07)	(0.00)	(0.49)	(0.00)	(0.62)
Education	-0.141	-0.146	-0.109	0.033	-0.117	-0.069
(1 = yes)	(0.00)	(0.00)	(0.00)	(0.99)	(0.01)	(0.10)
Health and social care	-0.12	-0.115	-0.127	-1.55	-0.111	-0.077
(1 = yes)	(0.01)	(0.01)	(0.00)	(0.36)	(0.01)	(0.06)
Other serv., private households	-0.097	-0.036	-0.121	-1.48	-0.108	0.003
(1=yes)	(0.15)	(0.59)	(0.01)	(0.57)	(0.09)	(0.96)
Temporary agency work	-0.036	-0.067	-0.014	0.646	0.023	0.099
(1 = yes)	(0.55)	(0.26)	(0.74)	(0.78)	(0.68)	(0.08)
Other sector	0.054	0.052	-0.077	-1.971	-0.039	0.019
(1 = yes)	(0.29)	(0.30)	(0.03)	(0.31)	(0.41)	(0.68)
Position last job (reference: professional)						
Helper job	0.060	0.077	0.001	-2.305	-0.005	-0.002
(1 = yes)	(0.11)	(0.04)	(0.98)	(0.11)	(0.88)	(0.96)
Complex specialist job	0.028	0.000	0.007	1.001	0.040	0.045
(1 = yes)	(0.41)	(1.00)	(0.76)	(0.45)	(0.22)	(0.17)
Highly complex job	0.054	0.002	0.035	4.596	0.035	0.049
(1 = yes)	(0.11)	(0.96)	(0.13)	(0.00)	(0.26)	(0.12)
Signup (reference: 1st quartile 2018)						
2nd quartile 2018	0.080	0.024	-0.043	2.794	-0.064	-0.012
(1 = yes)	(0.19)	(0.69)	(0.31)	(0.23)	(0.26)	(0.83)
3rd quartile 2018	0.103	0.073	0.002	3.707	-0.02	0.107
(1 = yes)	(0.07)	(0.20)	(0.95)	(0.09)	(0.71)	(0.05)
4th quartile 2018	0.024	-0.028	-0.006	6.133	-0.101	-0.011
(1 = yes)	(0.66)	(0.61)	(0.87)	(0.00)	(0.05)	(0.82)
1st quartile 2019	0.079	0.047	-0.013	3.731	-0.049	-0.019
(1 = yes)	(0.15)	(0.40)	(0.74)	(0.08)	(0.34)	(0.72)
2nd quartile 2019 or later	0.115	0.078	-0.034	4.675	-0.065	-0.002
(1 = yes)	(0.04)	(0.16)	(0.39)	(0.03)	(0.22)	(0.97)
Constant	0.434	0.346	0.194	10.952	0.514	0.405
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	2,327	2,327	2,327	2,327	2,327	2,327
R-squared	0.066	0.056	0.046	0.845	0.076	0.039

Source: GJSP and IEB (V16.00.01-202012).
Notes: Linear probability models / OLS.