

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

Gesine Stephan (Institut für Arbeitsmarkt- und Berufsforschung – IAB, Nuremberg)

Clemens Hetschko* (University of Leeds)

Julia Schmidtke (Institut für Arbeitsmarkt- und Berufsforschung – IAB, Nuremberg)

Michael Eid (Freie Universität Berlin)

Mario Lawes (Freie Universität Berlin)

September 2025

Acknowledgements. We thank the editor Michèle Belot as well as three anonymous reviewers for their highly valuable advice. Furthermore, we are grateful for comments by Ruben Bach, Michael Cooper, Simon Trenkle, as well as participants of the yearly meetings of RES (Birmingham, 2025), ISQoLS (Luxembourg, 2025), AIEL (Milan, 2025), EALE (Bergen, 2024), BeWell (Magdeburg, 2024), and SES (Glasgow, 2024). The DIM unit of IAB, in particular Stephan Griebemer, 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.socialscisceregistry.org/trials/2657>) in January 2018).

* *Corresponding author:* University of Leeds, Economics Department, Maurice Keyworth Building, Leeds LS2 9JT, United Kingdom, email-address: c.hetschko@leeds.ac.uk

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

September 2025

Abstract: We causally identify the effects of intense survey participation on key labor market outcomes, by randomly excluding individuals willing to sign up for a high-intensity survey with a focus on job search and well-being. Using administrative data, we find that, on average, survey participation had no effect on labor market outcomes during the year after signing up. We also demonstrate that an alternative selection-on-observables approach would have yield misleading results. This underscores the value of experiments in examining effects of survey participation.

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

JEL classifications: C83, C93, J63, J64

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 behavior, 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 result in changes of either actual behavior or reporting behavior (Chadi 2013, Eckman and Bach 2021, Bach 2021, Cernat and Keusch 2022). In both cases, survey participation has an impact on observed outcomes during later waves of a panel study, a phenomenon called panel conditioning. By means of a field experiment, we study Hawthorne effects and thus changes-in-behavior panel conditioning in job seekers who took part in a high-intensity survey.

Our contribution to the literature is at least threefold: First, we answer the question of whether participation in an intense monthly app-based survey affects subsequent labor market outcomes of job seekers, such as employment transitions (e.g., terminating a job, taking up a new job) and job quality (e.g., earnings). Much of the existing literature is related to the areas of voting, retirement savings, and health. It often confirms that participation indeed

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).

has a (context-dependent) impact on behavior (for overviews see, e.g., Bach 2021, Cernat and Keusch 2022).

Second, 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 are actually willing to participate in the survey prolongs the recruitment phase and requires additional resources. We therefore conduct a separate analysis employing a lower cost ‘selection-on-observables’ approach instead. The additional analysis compares the treatment group to individuals who were invited to take part in the survey but did not respond to the 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. While being less costly than the experiment, the selection-on-observables approach comes at the risk of confounding from unobserved differences between participants and non-participants.

Third, our analysis focuses on participation in a high-intensity panel survey involving detailed monthly measurements. Existing research on reactivity effects often stems from surveys that are relatively non-invasive. 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). By studying this type of survey, we contribute to a better understanding of reactivity effects in more demanding contexts. While yearly panel surveys are an indispensable source of information for economics research, higher-than-yearly frequency surveys become popular, too. Aside from the ‘German Job Search Panel’ we study here, the Dutch ‘Longitudinal Internet Studies for the Social Sciences’ effectively compartmentalizes its extensive yearly core study questions into shorter monthly modules, alongside additional custom surveys from various researchers (for an example using the data, see Archard et al. 2025). Some of the other traditional yearly household

panel surveys launched higher frequency spin-off surveys during the Covid-19 pandemic, such as the Understanding Society Covid-19 project with up to nine individual waves in 2020 and 2021 (e.g., Chaudhuri and Howley 2022).

1. Theoretical considerations and previous findings

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 beyond internal study validity if study participation was found to interfere with job search behavior 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. Moreover, job seekers may try harder to align their behavior with the social norm to work (e.g., Stutzer and Lalive 2004, Günther et al. 2025) when participating in a survey about job search makes their non-compliance with this norm more salient (Halpern-Manners et al. 2017). Overall, these are good reasons to expect survey participation to increase the probability of being employed in our context.

On the other hand, study participation as a contribution to the greater good could be seen as a way of compensating for a lack of job search effort (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). This would then actually increase the individual's probability of being unemployed and, hence, their dependency on public income support. In any case, 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 survey-related issues such as changes in reporting biases or panel attrition. When addressing the first challenge, the 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). 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 for study participation (Bach and Eckman 2019).

Regarding panel studies, actual Hawthorne effects (i.e., changes in behavior) need to be disentangled from changes in reporting behavior, 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. Alternatively, digital trace data may be used to investigate the impact of survey participation, however they come with substantial measurement challenges of their own, such as ‘non-response’ bias from the fact that the willingness to share digital trace data is non-random (Bähr et al. 2022, Cernat and Keusch 2022, Cernat et al. 2024).

We combine both gold standards to investigate if participation in a high-intensity panel survey affects the labor market outcomes of participants: First, we randomly assigned part of the individuals willing to participate in the German Job Search Panel (GJSP) to a *control group* excluded from participating in the survey. In contrast, *treatment group* individuals were allowed to continue participating. The GJSP was a high-intensity panel survey following the same people for up to two years (for details, see Hetschko et al. 2022). It used an

innovative survey app for frequent detailed measurements every month, including real-time assessments, a diary method and biomarker measurement (for details, see Section 2).

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

Three studies that also meet the two gold standards are Persson (2014), Crossley et al. (2017) and Zwane et al. (2011), none of which examined labor market outcomes. Persson (2014) showed that being randomly assigned to participating in a high-intensity election survey prior to the election increased turnout compared to participating after the election. Presumably 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. From administrative wealth data, they linked information on actual savings. They found 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 in a development context, Zwane et al. (2011) found that being surveyed about health increased the demand for water

treatment products and medical insurance, whereas being surveyed about borrowing behavior did not influence the demand for a microloan.

We are aware of only one study focusing on labor market effects of panel participation: Bach and Eckman (2019) use the random assignment of invitations to participate in the annual German Panel ‘Labor Market and Social Security’ within a larger population of eligible households as an instrumental variable for actual participation. They focus on welfare recipients and find that participation in the panel leads to increasing take-up of active labor market programs. While their instrumental variable strategy acknowledges the necessary assumptions, any such approach is open to speculation about whether the exclusion restriction holds.² Although their identification strategy is convincing, an experiment like ours eliminates even the most unlikely issues in this regard by design. Furthermore, the authors relate a specific population (welfare recipients) to a specific outcome (active labor market program participation). While our population is also specific (originally registered job seekers), our analyses cover a broader range of labor market outcomes, and survey participation was more invasive in that it occurred at a greater frequency (monthly instead of annually).

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). The 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 (e.g., Lawes et al. 2022, 2023, 2025, Schmidtke et al. 2024). Potential survey participants were drawn from not-yet unemployed job seekers who registered with the German Federal Employment Agency. To be eligible for

² For instance, people who are selected for the survey, but cannot be reached due to missing, outdated, or incorrect address data, might also be more difficult to engage in a training program.

unemployment benefits, job seekers must register three months before their employment ends, or within three days if they learn about the end of their employment later. Among these cases, a sizable fraction of workers actually entered unemployment, whereas a similarly large share remained employed (Stephan 2016). For instance, many workers register as job seekers as their fixed-term contract expires, but oftentimes 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 does not happen eventually. Among registered job seekers, people identified as part of an upcoming mass layoff were oversampled.³ The sample was restricted to individuals with German citizenship in order to avoid language issues with the survey questionnaires.

From November 2017 to May 2019, job seekers 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 included a random group assignment for the purpose of our field experiment. Around two thirds of the participants were randomly selected and allowed to further participate in the survey. In the following, we refer to these people as the *treatment group*. One third of eligible participants were excluded from further participation. These constitute our *control group*. Comparing the labor market outcomes of these two groups produces causal evidence about Hawthorne effects from GJSP participation.

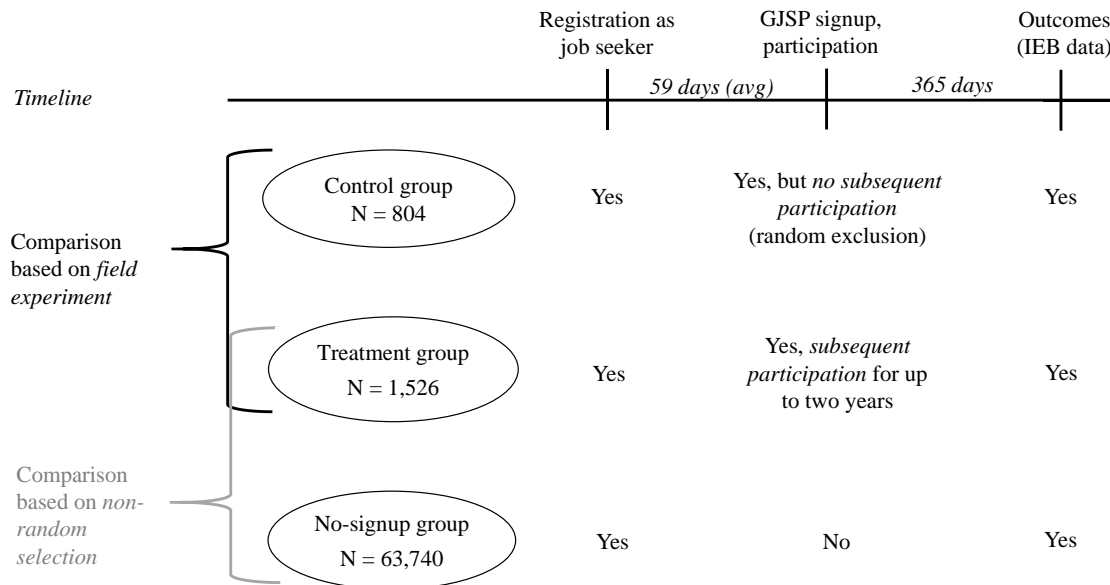
As mentioned above, additional time and other resources were needed to conduct the experiment, mostly because the sample filled up more slowly as people willing to participate were excluded. It is thus worthwhile to test whether a selection-on-observables approach simply comparing the labor market outcomes of people unwilling to partake with those of the survey

³ 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.

participants produces the same insights regarding Hawthorne effects. We therefore also separately compare a so-called *no-signup group* with the treatment group. These individuals were invited but did not participate in the entry survey. As Hetschko et al. (2022) show in their analysis of non-response to the GJSP, non-participation was non-random in respect to observable characteristics: High-skilled workers, young individuals, and women were more likely to sign up. Reassuringly for users of the GJSP, however, the average non-response bias across all characteristics examined by Hetschko et al. (2022) is rather low between 3% and 4%. We use the no-signup group in the analysis to find out about the scientific benefit of the costly field experiment and analyze whether controlling for observable characteristics would have led to the same conclusions regarding Hawthorne effects as the experimental design. A valuable alternative to this approach would have been to include a comparison with a randomly chosen group of registered job seekers that were not invited to take part in the survey. However, this was not possible for the larger part of our sample as all individuals registering as job seekers due to a mass layoff in Germany during the recruitment period were invited to participate in the GJSP.

Figure 1 provides an overview of the three groups and their roles in our study. After applying appropriate sample restrictions, our final sample comprises 1,526 persons in the treatment group, 804 persons in the control group, and 63,740 individuals in the no-signup group (see also Figure 1). In the Appendix, we document all sample restrictions in detail.

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



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 appointment data from the meeting scheduling software of the Federal Employment Agency. The IEB have been widely used in previous labor market research (e.g., Bossler et al. 2019, Dauth 2020, Bachmann et al. 2022). They contain 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 labor market programs administered by the Federal Employment Agency. Trainings taken up in our sample mostly (90%) comprise active labor market programs with a firm or private provider, and longer lasting further trainings. Appointment data cover information on scheduled, attended, and missed appointments of job seekers.

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 labor 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 using the mean number of days between the job seeker registration and survey signup observed in the treatment and control group.

We argue that taking part in the survey was a substantial burden on individuals in light of the monthly frequency of questionnaires, and the numerous questions to be answered each month. Monthly experience sampling (six measurements on one day every month, to be answered within 30 minutes) and quarterly day reconstructions were used to elicit momentary happiness and time use (Stone and Litcher-Kelly 2006, Kahneman et al. 2004).⁴ Cognitive well-being and mental health data were collected on a monthly basis 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 and send these in for the measurement of the stress hormone cortisol (for details, see Lawes et al. 2024).

Participants were also asked each month to indicate information about various socio-demographic characteristics, personality traits (three-monthly), coping resources, and their current labor market status. If unemployed, they were asked, for instance, about their reemployment prospects, reservation wage, and job search activities (e.g., ‘Have you actively looked for a job in the last four weeks?’). Employed individuals were asked about job characteristics, earnings, working hours and the likelihood of upcoming changes in their employment status (e.g., ‘How likely is it that the following changes [...] will occur within the next six months?’), followed by separate items for specific events such as ‘You look for a new position’, ‘You actually lose your job’, ‘You become self-employed.’ and so on’).

⁴ 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.

To spread out the burden of participation, different questionnaire modules would pop up in the survey app on up to eight days each month. The time needed to complete the daily surveys was about five minutes. On average, individuals responded to 150 items per survey month. On top of that, more time-consuming and burdensome measurements were carried out on selected days, especially experience sampling, the day reconstruction method and hair sampling.

While we argue that participation was quite intense, panel attrition works against treatment intensity in terms of Hawthorne effects. 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 weak treatment. 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. It is therefore reassuring that around 50% of the sample retained after the entry survey continued to participate for at least one year (i.e., they completed 13 monthly waves, see Hetschko et al. 2022). We examine the issues of treatment intensity and attrition in the course of our empirical analysis (see Section 4).

Table A.1 in the Online 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 social security contributions, in unemployment, and with benefit receipt). 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 in the Online Appendix 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, implying that any comparison between the treatment group and the no-sign up group is at risk of endogeneity bias. While the observed differences may partly be due to the considerably larger sample and, thus, enhanced statistical power, 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 (Hetschko et al. 2022).

3. Labor market outcomes and methods

We present findings for six outcome variables. With respect to duration outcomes, analyzing unconditional probabilities of transitions within certain durations is regarded the most appropriate method (e.g., Van den Berg et al. 2025). The randomization is compromised if the analysis is conditioned on survival at a specific time point, as the composition of survivors may vary within groups over time (Abbring and Van den Berg 2005). A competing risk analysis is thus unsuited for analyzing 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 labor market transitions and three outcomes that can be interpreted as job features or indicators of search effort. All outcomes are all measured until 360 days after signing up for the survey, as Hawthorne effects might take some time to arise and/or require repeated participation.

1. We first investigate if individuals transitioned out of regular *employment* during the 360 days after signing up, which we observe for half of all observations. ‘Regular employment’ is subject to social security contributions, thus excluding marginal employment, but including wage-subsidized employment. It may take place in a continuing or new employment relationship. Many job seekers search successfully for a new job when expecting to terminate an employment relationship without ever entering unemployment.⁵ We bridge gaps between two separate episodes of employment of up to 7 days to allow for short transitions between jobs.
2. As a natural counterpart, we analyze whether individuals entered *unemployment* after registering as a job seeker. This variable is not an exact mirror of employment exits. A substantial share of individuals transitions from employment into states other than unemployment.
3. Individuals who register as job seekers or are unemployed may take part in active labor 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 labor earnings, we impute a wage rate of zero. We analyze average earnings over a period of time as our sample was drawn from registered job

⁵ We do not analyze subsequent transitions out of unemployment as this would require us to condition on previously entering unemployment and into employment at the cost of compromising the randomization.

seekers. By conditioning on employment, we would lose the advantage of randomization.

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 individual search behavior, we examine if individuals have at least one *cancelled appointment* with the employment agency during the 360 days after 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 any non-participants and thus the control group.⁶

For each outcome, we estimate two specifications of linear probability models or OLS (for wages), respectively, to compare the treatment group with the control group (Section 4). First, we include only a dummy variable in the estimates for the treatment group, which constitutes a simple comparison of means. Second, we present the multivariate estimates controlling for a wide range of explaining variables. For a well-conducted field experiment, however, a comparison of means should already be sufficient to identify causal effects.

For further analyses of the treatment group and the control group, we include variables for the intensity of the treatment. In this context, we also discuss the possibility that attrition influences treatment effects via lowering treatment intensity.

Comparing the treatment group and the no-signup group (Section 5), any estimated effects of survey participation might rather reflect differences correlating with the willingness of

⁶ 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 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 Federal Employment Agency.

singing up than actual effects of survey participation. Observable characteristics are controlled for using 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.⁷

As we analyze several outcome variables, we account for multiple testing by conducting the Romano-Wolf 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 group and the control group across six outcomes in Table 1 as one group of hypotheses as the sample and the specification are the same.

4. Experimental results

Main findings

Table 1 presents our main results for the randomized experiment, displaying estimates for the treatment variable with and without covariates. It also informs about the respective mean values of the outcome variables for the reference group. In addition, Figure A.1 in the Online Appendix presents a coefficient plot for the estimates with covariates. Our first set of outcome variables focuses on transitions until 360 days after signing up. In the control group, around 53 percent of all individuals transitioned out of employment, 44 percent entered unemployment, and about 13 percent participated in a (short) training program. Table 1 shows

⁷ Entropy balancing works well in our sample with respect to the distribution of observable variables. Alternatively, we could have used propensity score matching (e.g., Dehejia and Wahba, 1999). It creates matched observations based on the estimated probability of receiving treatment. In additional analyses (not reported here), we found that this approach produces results similar to entropy balancing. Unlike the field experiment, both methods do not comprehensively tackle endogeneity issues arising from unobservable characteristics.

that all three transitions do not differ significantly between the treatment and the control group.⁸ 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. Compared to the control group mean, estimated relative effect sizes are generally of a small magnitude. This provides convincing evidence that participation in our demanding high-intensity survey did not have an impact on the labor market outcomes of participants.

Table 1 **Estimated effects of survey participation on labor market outcomes until 360 days after signup**

	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
Treatment group (1=yes)	-0.013 (0.56) {0.88}	0.014 (0.53) {0.88}	0.008 (0.60) {0.88}	3.766 (0.07) {0.29}	0.016 (0.43) {0.85}	-0.036 (0.08) {0.29}
Control variables	no	no	no	no	no	no
Treatment group (1=yes)	-0.012 (0.56) {0.93}	0.011 (0.62) {0.94}	0.004 (0.81) {0.94}	0.753 (0.36) {0.84}	0.008 (0.70) {0.94}	-0.035 (0.09) {0.35}
Control variables	yes	yes	yes	yes	yes	yes
Mean control group	0.532	0.438	0.134	108.531	0.325	0.336

Source: GJSP and IEB (V16.00.01-202012). Observations: 1,526 in the treatment group, 804 in the control group.
Notes: The table displays coefficients, uncorrected p -values (in parentheses) and multiple-hypothesis corrected p -values (in braced brackets) of linear probability models / OLS. List of control variables: See Table A.1 in the Online-Appendix for descriptive statistics of the covariates and Table A.3 for full regression results. Corrected p -values are computed using the Romano-Wolf correction for multiple-hypothesis-testing (Clarke et al. 2021).

Even if we find no effects for labor market transitions, survey participation might still have an impact on job quality. Within 360 days after 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 daily earnings in the treatment group are around 4 euros higher than in the control group, but this difference is not statistically significant.

⁸ Note that under standard assumptions (power = 0.80, significance level = 0.05), for our randomized sample (N = 2,330) with a treatment share of 0.65, the minimum detectable effect for a dummy variable with a mean of 0.134 (the control group mean for training participation, which has the smallest mean value among our dependent variables) would be 0.042, or 4.2 percentage points.

Most outcomes discussed up to here 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 labor market conditions. In the following, we therefore examine two more direct measures of job search efforts, namely job take-up in a different municipality and attending appointments with the employment agency.

First, we investigate if a job seeker took up a job in a different municipality within 360 days of signing up for the survey. Indeed, around one third of job seekers started working in a different municipality. However, we find no statistically significant differences between the treatment and the control group. Second, we check whether the individual job seeker had a scheduled appointment at their local employment agency and whether that meeting took place.⁹ Canceled appointments are mostly due to the job seeker not showing up. As the duration of both job search and of potential unemployment varies across individuals, we analyze whether at least one appointment scheduled with the local employment agency did not take place within the 360 days after signing up for the survey. The share of individuals with at least one missed appointment was around one third. We find, however, no statistically significant differences between the treatment and the control group.

Treatment intensity and attrition

To investigate issues of treatment intensity and attrition, we revisit the effect of survey participation for parts of the treatment group who were intensively treated. We interact survey participation with two indicators of treatment intensity: continued participation until at least

⁹ These meetings take place between the job seeker and a staff member responsible for their case. Employment agencies offer an appointment for an early meeting soon after registration as a yet-employed job seeker. While early meetings do not prevent unemployment, the literature shows they significantly accelerate subsequent job finding (Homrighausen and Oberfichtner 2024, Rosholm 2014, Schiprowski 2020).

month 7 and participation in the cortisol study, which involved sending in strands of hair to receive an objective stress measure. For the control group, both treatment intensity dummies were assigned a value of zero. As a result of non-random nonresponse, the findings based on this indicator of treatment intensity should be cautiously interpreted. Only 26% of the sample were eligible (e.g., sufficient hair length) and willing to partake in the hair sampling (Lawes et al. 2024).

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 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 of treatment intensity.

Table 2 Estimated effects of participation intensity on labor market outcomes until 360 days after signup

Treatment vs. control group	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. Continued survey participation						
Treatment group (1=yes)	-0.010 (0.68) {0.97}	0.025 (0.32) {0.86}	0.005 (0.78) {0.97}	0.917 (0.35) {0.88}	-0.001 (0.96) {0.97}	-0.022 (0.37) {0.88}
Participated at least until month 7 (1=yes)	-0.004 (0.88) {1.00}	-0.027 (0.29) {0.86}	-0.002 (0.89) {1.00}	-0.305 (0.75) {1.00}	0.017 (0.48) {0.92}	-0.025 (0.30) {0.86}
Control variables	yes	yes	yes	yes	yes	yes
II. Cortisol study participation						
Treatment group (1=yes)	-0.015 (0.51) {0.90}	0.011 (0.63) {0.95}	0.003 (0.86) {0.98}	0.749 (0.39) {0.89}	-0.004 (0.85) {0.98}	-0.034 (0.11) {0.45}
Participated in cortisol study (1=yes)	0.010 (0.74) {1.00}	-0.001 (0.98) {1.00}	0.003 (0.90) {1.00}	0.014 (0.99) {1.00}	0.045 (0.10) {0.46}	-0.002 (0.95) {1.00}
Control variables	yes	yes	yes	yes	yes	yes

Source: GJSP and IEB (V16.00.01-202012). Observations: 1,526 in the treatment group, 804 in the control group.
Notes: The table displays coefficients, uncorrected *p*-values (in parentheses) and multiple-hypothesis corrected *p*-values (in braced brackets) of linear probability models / OLS. List of control variables: See Table A.1 in the Online Appendix. Corrected *p*-values are computed separately for Panels I and II, using the Romano-Wolf correction for multiple-hypothesis-testing (Clarke et al. 2021).

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 randomization 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 continuously participate, however the effect is weakly significant only. As we show further below, females do not differ when it comes to the treatment effects. Overall, there is little evidence to suggest that our results are confounded by attrition.

Heterogeneous effects

We first conduct a heterogeneity analysis based on theoretical considerations: For our experimental sample, 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, that is, they were re-employed by a previous employer (e.g., seasonal workers). 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., Kassenboehmer and Haisken-DeNew, Schmieder et al. 2023). A gender-specific analysis seems appropriate as the labor market behavior 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 a recall have a smaller incentive to exert search effort.

The results are presented in Table 3, controlling for the full set of covariates and correcting for multiple-hypothesis-testing. We find no significant interactions between survey participation and female gender, being on a temporary contract, or having been recalled to a previous job. 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 they generally enter unemployment later than those who were dismissed for other reasons.

Table 3 Heterogeneous effects of participation intensity on labor market outcomes until 360 days after signup

Treatment vs. control group	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
Treatment group (1=yes)	0.044 (0.37) {0.88}	0.045 (0.35) {0.88}	0.018 (0.60) {0.89}	3.981 (0.03) {0.16}	0.014 (0.76) {0.89}	-0.031 (0.50) {0.89}
Treatment group * mass layoff	-0.095 (0.03) {0.16}	-0.113 (0.01) {0.06}	-0.015 (0.64) {0.94}	-0.078 (0.96) {0.96}	-0.042 (0.32) {0.78}	-0.014 (0.75) {0.95}
Treatment group * female	0.045 (0.31) {0.71}	0.078 (0.07) {0.34}	-0.029 (0.34) {0.71}	-2.398 (0.15) {0.53}	0.028 (0.50) {0.71}	0.043 (0.30) {0.71}
Treatment group * temporary contract	-0.023 (0.61) {0.96}	0.004 (0.93) {0.99}	0.005 (0.88) {0.99}	-1.844 (0.29) {0.82}	0.007 (0.87) {0.99}	-0.031 (0.47) {0.93}
Treatment group * recall during last 5 years	-0.026 (0.62) {0.94}	-0.033 (0.52) {0.94}	0.032 (0.38) {0.83}	-3.148 (0.11) {0.41}	0.003 (0.95) {1.00}	0.002 (0.96) {1.00}
Mass layoff (1=yes)	0.062 (0.10) {0.41}	0.078 (0.04) {0.18}	0.004 (0.89) {1.00}	0.406 (0.78) {0.99}	0.007 (0.85) {1.00}	0.011 (0.75) {1.00}
Female (1=yes)	-0.053 (0.15) {0.34}	-0.086 (0.02) {0.10}	0.013 (0.61) {0.82}	-0.789 (0.57) {0.82}	-0.071 (0.04) {0.16}	-0.062 (0.07) {0.24}
Temporary contract (1=yes)	-0.041 (0.31) {0.45}	-0.021 (0.60) {0.53}	-0.053 (0.06) {0.21}	1.845 (0.23) {0.45}	-0.076 (0.04) {0.17}	0.055 (0.14) {0.35}
Recall during last 5 years (1=yes)	0.167 (0.00) {0.00}	0.049 (0.23) {0.55}	-0.052 (0.08) {0.25}	-0.883 (0.58) {0.81}	-0.11 (0.01) {0.02}	-0.011 (0.78) {0.81}
Control variables	yes	yes	yes	yes	yes	yes

Source: GJSP and IEB (V16.00.01-202012). Observations: 1,526 in the treatment group, 804 in the control group.

Notes: The table displays coefficients, uncorrected *p*-values (in parentheses) and multiple-hypothesis corrected *p*-values (in braced brackets) of linear probability models / OLS. List of further control variables: See Table A.1 in the Online Appendix. Corrected *p*-values are computed using the Romano-Wolf correction for multiple-hypothesis-testing (Clarke et al. 2021)

Second, we estimate causal forests to identify at a more general level whether treatment effect heterogeneity is present in any of our outcomes (e.g., Wager and Athey 2018).¹⁰ The estimated average treatment effects of the causal forests are practically identical to those obtained from our regression analysis. The conditional average treatment effect (CATE) represents the individual treatment effect conditional on covariates. Based on a median split, we compute a binary variable for belonging to an either low or high CATE group. Next, we interact the treatment indicator with this CATE variable to analyze whether treatment effects differ significantly between the two groups. The resulting interaction effect is statistically significant for two of our six outcomes only, training participation and taking up a job in another municipality, suggesting some degree of treatment effect heterogeneity in these two outcomes. To analyze whether any of our covariates might explain the heterogeneity, we interact all control variables with our treatment indicator. The results are displayed in Table A.4 in the Online-Appendix (for all outcomes). When correcting for multiple-hypothesis testing none of the initially six statistically significant interaction effects remains significant. This implies that variables not present in the dataset explain the treatment effect heterogeneity in training participation and taking up a job in a different municipality. Overall, however, we find little evidence of treatment heterogeneity across subgroups. At least for most of our outcomes both regression analysis and causal forest imply zero effects that do not originate from mutually canceling subgroup effects.

5. Comparison with the no-signup group

As outlined above, we are also interested in the question if our randomized experiment was worth the effort. Therefore, we also compare the treatment group with a no-signup group.

¹⁰ We use the R package *grf* (Tibshirani et al. 2025) for estimating a causal forest for each of our six outcome variables. We apply cross-validation with five folds and loop over three random seeds to decrease potential dependence of our results on a specific random seed. We estimate 5,000 trees in each iteration using honest trees and tuning all model parameters by default.

Table 4 presents findings without (Panel I) and with (Panel II) entropy balancing, and Figure A.1 in the Appendix displays coefficient plots for the full estimates. Without entropy balancing, and including only the treatment variable but no further control variables, we find no significant differences in transitions and the number of cancelled appointments (at least, once we correct for multiple-hypothesis-testing).

Table 4 Comparison of labor market outcomes until 360 days after signup between survey participants and the no-signup group

	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. Without entropy balancing						
Treatment group (1=yes)	0.018 (0.17) {0.35}	0.008 (0.53) {0.53}	0.017 (0.04) {0.16}	15.795 (0.00) {0.00}	0.031 (0.01) {0.06}	-0.015 (0.21) {0.37}
Control variables	no	no	no	no	no	no
Treatment group (1=yes)	0.008 (0.54) {0.85}	0.004 (0.74) {0.93}	0.024 (0.01) {0.05}	1.029 (0.02) {0.09}	0.023 (0.05) {0.16}	-0.001 (0.91) {0.93}
Control variables	yes	yes	yes	yes	yes	yes
Mean no-signup group	0.502	0.443	0.125	96.502	0.31	0.315
II. With entropy balancing						
Treatment group (1=yes)	0.007 (0.61) {0.88}	0.003 (0.82) {0.98}	0.024 (0.01) {0.05}	1.007 (0.42) {0.88}	0.026 (0.04) {0.19}	0.001 (0.94) {0.98}
Control variables	no	no	no	no	no	no
Treatment group (1=yes)	0.007 (0.59) {0.93}	0.003 (0.81) {0.96}	0.024 (0.01) {0.04}	0.943 (0.05) {0.18}	0.025 (0.03) {0.15}	0.001 (0.93) {0.96}
Control variables	yes	yes	yes	yes	yes	yes
Mean no-signup group	0.513	0.448	0.118	111.291	0.315	0.299

Source: GJSP and IEB (V16.00.01-202012). Observations: 1,526 in the treatment group, 63,740 in the no-signup group. Notes: The table displays coefficients, uncorrected p -values (in parentheses) and multiple-hypothesis corrected p -values (in braced brackets) of linear probability models / OLS. List of control variables: See Table A.1 in the Online Appendix. Corrected p -values are computed separately for Panels I to II, using the Romano-Wolf correction for multiple-hypothesis-testing (Clarke et al. 2021).

Controlling for observable attributes of both groups, however, the results suggest a significantly positive relationship of survey participation and 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. We obtain the same results if we use entropy balancing to achieve similar distributions of observable characteristics in the no-signup group and the treatment group. Taken together with

the experimental analysis in which we found no such effects, this implies that at least some unobserved differences between the treatment group and the no-signup group remain after controlling for observable characteristics, and that these unobserved differences are correlated with the propensity of participating in training.

Finally, we find significant differences in daily earnings and work in another municipality, if we compare the treatment group with the no-signup group and do not control for further covariates (upper part Panel I). These differences, 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.

6. Conclusions

We investigated if participation in an intensive app-based survey on job search and well-being had an impact on labor market outcomes within a year of signing up for the survey. To this end, we combined two gold standards of research into Hawthorne effects: First, we conducted a field experiment, randomly excluding one third of individuals willing to partake in a survey from participation for use as a control group of actual survey participants. Second, we merged information on survey participation with administrative data on labor market outcomes, ruling out that our results are in any way related to reporting bias.

Our most important finding is that participation in the survey, on average, had no impact on any of the investigated labor market transitions of initially employed job seekers, namely out of employment, into unemployment, and into subsidized (short) training. We also found no effects of survey participation on daily wages, taking up a job in a different municipality, and cancelled appointments with the employment agency. There was also little evidence for effect heterogeneity across subgroups. A causal forest analysis implied some heterogeneous

effects on the outcomes of training participation and taking up a job in a different municipality only. This heterogeneity appeared unrelated with the variables in our dataset and thus constitutes a topic for future research.

In addition, we showed that even controlling for a wide range of observable characteristics and correcting for multiple-hypothesis-testing, a comparison of survey participants with individuals not signing up for the survey would have led to misleading conclusions. Regression results showed that survey participants statistically and significantly more often take up subsidized (short) training if compared to the no-signup group. Thus, there was some remaining selection into survey participation based on unobservable characteristics, creating a false sense of an impact of survey participation on training participation in a selection-on-observables setting. 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 other resources to fill up our sample.

An important question is how generalizable our main findings from the experimental study are. Previous research found that outcomes such as saving (Crossley et al. 2017) or voting (Persson 2014) reacted to survey participation, while we found no such effects for labor market outcomes. Thus, the specific study context clearly matters. One reason for the lack of measurable reactivity effects in our experiment may be that most of our outcomes (e.g., having a job at a certain point in time) are less controllable by the survey respondent than those for which the previous literature has found effects (e.g., turning up at a polling station). The fact that our outcomes are partly influenced by characteristics reflecting a person's employability (e.g., previous work experience, labor demand) may weaken the link between behavioral change triggered by survey participation and realized outcomes.

Other aspects of the survey we examined likely enhance the generalizability of our results, at least when it comes to labor market outcomes. Participation in the GJSP was highly intense, due to monthly measurements with modules appearing on respondents' smartphones over several days of each month, including real-time measurement and diary methods, as well as hair sampling. A logical assumption regarding survey participation effects on labor market outcomes is that any such impact would increase with treatment intensity, defined by the extent and frequency of being surveyed. Yet our experimental study found no such effects, suggesting that most other surveys relevant to labor market research, such as less intense yearly household surveys, are also unlikely to influence real-world labor market outcomes.

On the contrary, more targeted surveys, even if less intense than the GJPS, might impact behavior more significantly. Unlike the GJPS with its broad scope (job search, employment, well-being, health), a more targeted survey may focus respondents' attention on certain areas, in particular if these are of relevance for the specific population. Bach and Eckman (2019), for instance, found survey effects on participation in active labor market programs among welfare recipients. Our population of registered job seekers was perhaps less likely to show these effects because they were not the primary target of such programs. A significant share of these job seekers did not enter unemployment at all, and those who became unemployed typically received unemployment insurance benefits. In contrast, Bach and Eckman (2019) analyze welfare benefit recipients, who were overwhelmingly long-term unemployed and thus more strongly targeted by active labor market policies.

Notwithstanding the potential caveat of context-dependency, our findings provide good news for survey researchers especially in the area of labor economics. The lack of reactivity effects speaks to the internal validity of research results obtained from analyzing survey data, even in cases where participation is frequent and burdensome. Further research should aim

to obtain a more complete picture of the circumstances under which reactivity occurs. For instance, future studies could study different populations of survey participants, different countries, and different labor market conditions.

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.

Achard, P., Albrecht, S., Ghidoni, R., Cettolin, E., & Suetens, S. (2025). Local exposure to refugees changed attitudes to ethnic minorities in the Netherlands. *Economic Journal*, 135(667), 808-837.

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.

Bachmann, R., Demir, G., & Frings, H. (2022). Labor market polarization, job tasks, and monopsony power. *Journal of Human Resources*, 57(S), S11-S49.

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.
- Bossler, M., Mosthaf, A., & Schank, T. (2020). Are Female Managers More Likely to Hire More Female Managers? Evidence from Germany. *ILR Review*, 73(3), 676-704.
- Cernat, A., & Keusch, F. (2022). Do surveys change behaviour? Insights from digital trace data. *International Journal of Social Research Methodology*, 25(1), 79–90.
- Cernat, A., Keusch, F., Bach, R.L. and Pankowska, P.K., (2024, forthcoming). Estimating measurement quality in digital trace data and surveys using the MultiTrait MultiMethod model. *Social Science Computer Review*.
- Chaudhuri, K., & Howley, P. (2022). The impact of Covid-19 vaccination for mental health. *European Economic Review*, 150, 104293.
- 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? *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.
- Dauth, C. (2020). Regional Discontinuities and the Effectiveness of Further Training Subsidies for Low-Skilled Employees. *ILR Review*, 73(5), 1147-1184.
- Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448), 1053-1062.
- 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.
- Eckman, S., & Bach, R. (2021). Panel conditioning in the us consumer expenditure survey. *Journal of Official Statistics*, 37(1), 53-69.
- 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. (2025). Socialism, identity and the well-being of unemployed women. *Labour Economics*, 102752.
- 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/7jazzr>.
- Homrighausen, P., & Oberfichtner, M. (2024). Do caseworker meetings prevent unemployment? Evidence from a field experiment. IZA Discussion Paper 16923.
- 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. (2024). Collecting hair samples in online panel surveys: Participation rates, selective participation, and effects on attrition. *Survey Research Methods*, 18(2), 167-185

Lawes, M., Hetschko, C., Schöb, R., Stephan, G., & Eid, M. (2025). 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*, 39(1), 24-45.

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.

- Rosholm, M. (2014). Do case workers help the unemployed? IZA World of Labor.
- 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.
- Schiprowski, A. (2020). The role of caseworkers in unemployment insurance: Evidence from unplanned absences. *Journal of Labor Economics*, 38(4), 1189-1225.
- Schmidtke, J., Hetschko, C., Schöb, R., Stephan, G., Eid, M., & Lawes, M. (2024). Does worker well-being adapt to a pandemic? An event study based on high-frequency panel data. *Review of Income and Wealth*, 70(3), 840–861.
- 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.
- Tibshirani, J., Athey, S., Friedberg, R., Hadad, V., Hirshberg, D., Miner, L., Sverdrup, E., Wager, S., Wright, M. (2025). Package ‘grf’. Generalized Random Forests. Version 2.4.0. <https://github.com/grf-labs/grf>

- Van den Berg, G. J., Hofmann, B., Stephan, G., Uhlendorff, A. (2025). Mandatory integration agreements for unemployed job-seekers: A randomized controlled field experiment in Germany. *International Economic Review*, 66(1), 79-105.
- Van Landeghem, B. (2014). A test based on panel refreshments for panel conditioning in stated utility measures. *Economics Letters*, 124(2), 236-238.
- Wager, S., Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242.
- 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.

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).¹¹ 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.¹² 940 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 disproportionately 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

¹¹ 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.

¹² 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).

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 1,526 persons in the treatment group, 804 persons in the control group, and 63,740 individuals in the no-signup group (see also Figure 1).