

van den Berg, Gerard J.; Hofmann, Barbara; Stephan, Gesine; Uhlendorff, Arne

Article — Published Version

Mandatory integration agreements for unemployed job seekers: a randomized controlled field experiment in Germany

International Economic Review

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: van den Berg, Gerard J.; Hofmann, Barbara; Stephan, Gesine; Uhlendorff, Arne (2024) : Mandatory integration agreements for unemployed job seekers: a randomized controlled field experiment in Germany, International Economic Review, ISSN 1468-2354, Wiley, Hoboken, NJ, Vol. 66, Iss. 1, pp. 79-105,
<https://doi.org/10.1111/iere.12745>

This Version is available at:

<https://hdl.handle.net/10419/323782>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

MANDATORY INTEGRATION AGREEMENTS FOR UNEMPLOYED JOB SEEKERS: A RANDOMIZED CONTROLLED FIELD EXPERIMENT IN GERMANY*

BY GERARD J. VAN DEN BERG, BARBARA HOFMANN, GESINE STEPHAN, AND ARNE UHLENDORFF

University of Groningen, University Medical Center Groningen, Netherlands and IFAU Uppsala, Sweden; FEA Nuremberg, Germany; Institute for Employment Research (IAB Nürnberg), Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany; CNRS, CREST, and IAB Nuremberg, 5, Av Henry Le Chatelier, Palaiseau 91120, France

Integration agreements (IAs) are contracts between the employment agency and the unemployed, nudging the latter to comply with rules on search behavior. We designed and implemented a randomized controlled trial involving thousands of newly unemployed workers, randomizing at the individual level both the timing of the IA and whether it is announced in advance. Administrative records provide outcomes. Novel theoretical and methodological insights provide tools to detect anticipation and suggest estimation by individual baseline employability. The positive effect on entering employment is driven by individuals with adverse prospects. For them, early IA increase reemployment within a year from 53% to 61%.

1. INTRODUCTION

During the past decades, a view has emerged that Active Labor Market Programs (ALMP) are on average not very effective in bringing unemployed individuals back to work. Specifically, average reemployment effects of participation in training and workfare are often rather low, whereas the effectiveness of job search assistance and monitoring varies with the setting at hand and is typically low for groups with relatively bleak labor market prospects. Crépon and van den Berg (2016) and Card et al. (2018) provide recent overviews, but discouraging findings were already documented and summarized as early as Heckman et al. (1999). The evidence is of concern even in labor markets with favorable conditions, as unemployment may drive individuals out of the regular labor market and, indeed, may lead them to drift away from mainstream society. This has led to a search for different tools to support the placement of the unemployed.

In this article, we evaluate one such policy, called mandatory integration agreements (IAs). An IA is a written contract that stipulates rights and obligations of an unemployment insurance (UI) recipient. The signing of this contract takes place upon entry into UI, at the end of the first meeting of the UI recipient, and his/her caseworker in the employment agency.¹ Both

*Manuscript received July 2022; revised May 2024.

We are grateful to the Editor Petra Todd and two anonymous Referees as well as to Bart Cockx, Bruno Crépon, Jeffrey Smith, Johan Vikström, Michael Visser and participants in conferences in Ghent, Seville, Mannheim, Bonn, Bristol, Augsburg, Groningen and Paris, seminars in Potsdam, Sheffield, Paris (CREST and DARES), Lucca and St Gallen, and a J-PAL summer school, for helpful comments. We thank Christian Rauch for enabling the project, Susanne Koch and Carina Schwarz for their cooperation in the set-up and implementation of the experiment, the “ProIAB” institutional experts of the IAB for their support in implementing the RCT on-site, and the IAB-DIM (in particular Ali Athmani, Steffen Kaimer and Markus Köhler) for their technical and data-related support. The RCT is registered in the AEA RCT registry (0006368); note that the RCT took place before this registry was formally established. We gratefully acknowledge support of DFG SFB 884. Arne Uhlenndorff is grateful to Investissements d’Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047) for financial support. Please address correspondence to: Arne Uhlenndorff, CNRS, CREST, IAB Nuremberg. E-mail: arne.uhlenndorff@gmail.com.

¹ For ease of exposition, we refer to “being a UI recipient” as “being unemployed.” In Section 2, we discuss subtle differences.

the UI recipient and the caseworker should sign the IA. Its content is based on a template of textual building blocks that may slightly vary across occupation and family status but in practice the template is rather uniformly specified. The template reflects existing rules and laws (see also Schütz et al., 2011; and Boockmann et al., 2013, for descriptions of the IA; see also below).

Since the IA does not impose constraints that tighten the existing rules, one could argue that it is not being perceived as a monitoring device, at least as long as the unemployed individual is aware of the existing rules. Instead, the design and phrasing of the IA suggest a “nudge” character of the policy. Signing the IA, with its apparently symmetric design with rights and obligations and with its space for two signatures, may be viewed as a sort of ritual that may increase the commitment on both sides and foster a cooperative bond between the unemployed and the caseworker, effectively reducing the disutility of search as perceived by the unemployed. However, the content of the IA consists mostly of a list of obligations on the job search activities on the part of the unemployed worker, such as a minimum number of job applications per time unit. Even some of the stated rights of the unemployed can be seen as veiled threats to comply. Moreover, the caseworker may impose the IA unilaterally if the unemployed refuses to sign, and the resulting contract is legally binding. The punishment for noncompliance with aspects of IA is one-sided and involves UI benefits reductions. With all this in mind, and given the self-reported assessments of the IA by surveyed workers and caseworkers (see Section 2), it is more accurate to view the IA as a combination of a nudge and a refresher on obligations including a reminder of monitoring and potential punishments. As such, the IA may have effects similar to monitoring, where the effects may be stronger because of the nudging that may make the IA a more comprehensive experience than the alternative of a simple confrontation with a list of obligations.

We employ a randomized controlled trial (RCT) with randomization at the individual level to evaluate the IA. Specifically, we randomize two aspects of the policy: the timing of the IA and the advance notification of the IA. Randomization takes place upon entry into unemployment. The timing of the IA is randomized over possible elapsed times from entry into unemployment until the IA. One treatment arm involves the IA in the first month, one involves the IA at three months, and one at six months. In addition, we randomize whether those assigned to receive the IA at three months also receive an advance notification of the timing of the future IA at three months, to be received upon entry into unemployment. In total, these constitute four possible treatment statuses, each with a 25% assignment probability. The RCT was carried out in five local labor market regions in Germany. These were chosen for reasons of representativeness but also because they are large, because no ALMP pilots or other evaluations were held there, and no reorganizations took place in the local employment agencies at the time. It was promised to the agencies that their performance ratings would not be affected by the RCT.

We use a number of data sources. First, we observe the output of the randomization tool. Second, administrative data on UI recipients provide daily observations on outcomes: ALMP participation (including IA), meetings, covariates, employment spells, and past labor market outcomes. Third, we held a survey of caseworkers working in the agencies that participate in the RCT, one month before the RCT began. Fourth, we carried out a survey of UI recipients around two months after entry into UI. The nonresponse in the UI recipients' survey was sizeable. For data protection reasons, the survey data can only be linked to the administrative data if the respondent agrees. In particular, most of the caseworkers did not allow merging of their own responses to records of their clients. For these reasons, the survey data are of limited use. We merely use them to informally gauge workers' and caseworkers' perceptions of the IA.

Our RCT is the first causal evaluation of the IA policy. Also, it is the first large-scale RCT of ALMP in Germany with randomization at the individual level. Note that the combination of monitoring with nudging makes our evaluation potentially relevant for other policies that

combine these components, such as devices to discourage tax avoidance.² In addition, some other OECD countries have recently implemented what could be called weak versions of the IA policy (Immervoll and Knotz, 2018; Knotz, 2018), usually without the formal contract-signing ceremony and without threats of enforcement. In our view, an evaluation of the full-blown IA in Germany, with its strong legalistic tradition and adherence to the law, provides an interesting benchmark. We discuss policy differences in more detail in Subsection 2.2.

The comparison of those who are notified about a future IA at three months to those who are not is an innovative feature of our study design. This feature connects our article to the literature on anticipation of future treatments (see, e.g., Black et al., 2003; and van den Berg et al., 2009).³ Using a search-theoretical framework, we show that the two treatment arms lead to an observationally distinct difference in the reemployment rate around the three-month threshold. At first sight, it may seem that inference on the latter is hampered by the challenge that randomization is lost when conditioning on survival until close to three months. However, in the article we develop a novel method to detect qualitative features of the reemployment rate that are informative on the presence of anticipation of the treatment at three months and that are preserved if randomization is lost. Clearly, this has wider relevance for the evaluation of anticipatory effects of future events.⁴

To investigate heterogeneity of effects, we divide the population of unemployed into two groups based on their predicted median unemployment duration until reemployment. Predictions are based on an inflow sample into unemployment from the year before our experiment, conditioning on individual labor market histories and characteristics. We show that local labor market conditions in these years are stable. We split the sample into individuals with a high (above six months) and a low (below six months) predicted median duration and perform sensitivity analyses with respect to this threshold value. We do not use in-sample observations to quantify the prediction model in order to avoid overfitting and the related risk of biased treatment effects (Abadie et al., 2018).

Interestingly, our findings already led to a policy change in the use of IAs by the German Federal Employment Agency (FEA). Specifically, by now, individuals who are regarded as having favorable labor market prospects are not obliged anymore to undergo an IA during the first three months of unemployment.

² From a sociological-institutional perspective, IAs can be seen as an example of new public management strategies or new public contractualism with reconstructed citizens—in our case, job seekers—as customers (O’Flynn, 2007). In this view, contracts that define requirements, monitoring, and incentives constitute the legitimate relationship between the state as the principal and the job seeker as the agent.

³ Effects of advance announcements and notifications of future treatments are hard to identify because they are often not observed and they may obliterate the very treatment they announce, if they cause an exit from the state that is the eligibility state for the treatment. Nonexperimental studies have relied on policy discontinuities (Blundell et al., 2004; De Giorgi, 2005; van den Berg et al., 2020) or on self-reported assessments of the likelihood of a treatment in the near future under unconfoundedness (van den Berg et al., 2009) or on register data with observed advance announcements in a timing-of-events model setting (Lalive et al., 2005; Crépon et al., 2018). RCTs are uniquely equipped to study anticipation effects because advance announcements are predetermined by the study design. Michalopoulos et al. (2005) find a small increase in the share of individuals who anticipate access to in-work benefits by extending their length of stay on welfare in order to become eligible for those benefits, in the context of the Canadian Self-Sufficiency Project RCT. Büttner (2008) estimates effects of the announcement of participation in a future job search assistance program. However, sample sizes are in the low 100s and he resorts to propensity score methods to deal with implementation issues.

⁴ As indicated above, there is a large literature on effects of job search assistance, counseling and monitoring of unemployed job seekers. Some of this has studied policy measures that cannot be straightforwardly assigned to one of these standard ALMP categories, and in some cases the estimated effects shed a favorable light on their effectiveness. In general, caseworker meetings are deemed to speed up reemployment (see, e.g., Schiprowski, 2020, for a recent example). Altmann et al. (2018) and Belot et al. (2019) find that providing information to unemployed job seekers about job search and labor market opportunities may affect their job search behavior, leading to better employment prospects. Altmann et al. (2018) show that this is especially relevant for job seekers with a low predicted probability of leaving unemployment for a job. Our study extends this literature by providing additional evidence on non-standard policy measures. Similarly, our study extends the large literature on effects of monitoring schemes (for recent evidence, see, e.g., Lachowska et al., 2015) by examining a new tool combining monitoring with nudging.

The outline of the article is as follows: Section 2 describes the German UI benefit system and IAs. Sections 3 and 4 discuss the setup of the experiment and the data, respectively. Methodological considerations and novel methodological contributions are in Section 5. The empirical results are presented in Section 6. Section 7 concludes.

2. INSTITUTIONAL BACKGROUND

2.1. *UI Benefits.* The German unemployment compensation system has two pillars. The first is UI. As a norm, upon inflow into unemployment, UI eligibility requires that individuals have been working and paying social security contributions for at least 12 months within the period of 24 months immediately prior to unemployment (30 months since 2020). UI benefit recipients have to be registered as unemployed at the FEA. The UI entitlement duration depends on the duration of the prior employment period and the age of the recipient. The highest possible entitlement duration for individuals below 50 years is 12 months. This increases for older individuals, up to 24 months for those aged above 58 if they were employed for at least 48 months in the five years prior to unemployment. The replacement ratio is about 67% for individuals with dependent children and about 60% for those without, with a benefits level cap that is binding for only a small percentage of newly unemployed.

After expiration of UI, unemployment compensation is reduced to unemployment assistance or “welfare.” This is the second pillar of the system. Welfare is tax-financed and means-tested, and the level depends on household composition but not on former earnings. In 2012, it equaled around 345 euro per month with supplementary accommodation costs as well as support in case of specific needs. Recipients have to register and receive placement services in job centers that are partly administered by the FEA and partly by municipalities. In the article, we restrict attention to UI benefit recipients.

2.2. *Integration Agreements.* IAs were introduced as a policy in Germany in 2002.⁵ Appendix 2 provides an actual example of an IA for an unemployed physiotherapist, with a slightly abridged English translation. Most of its content is uniform across all IAs but a few features may vary across occupations. The latter applies in particular to the geographical range of the job search (here: nationwide), the minimum number of applications per month, the maximum time allowed for submitting a list of qualifications to successfully exert one’s occupation (typically one week), and the time until the next meeting (here: two months). As already discussed in Section 1, most of the IA text is about UI recipients’ obligations, and even some of the text on the recipients’ rights can be interpreted as a reminder of obligations or as a veiled threat in case of noncompliance (e.g., that the agency promises to make a phone call if it identifies an appropriate vacancy and in some cases may immediately send an actual job offer).

As noted in Section 1, the IA is signed at the end of the first meeting of the UI recipient and his/her caseworker. Before that, the meeting covers formalities such as entering information about the client into the computer system of the FEA, and a discussion of plans for job search and future participation in ALMP. This includes the information on occupation, qualifications, and household status that may affect the few open details of the IA to be signed. According to our survey among caseworkers (see Subsection 2.3), the first meeting usually takes about 50 minutes, and of these, about 15 minutes are used for the IA. Regarding the timing of the first meeting, we should point out that individuals are required to register as a job seeker three months before unemployment entry or—if they do not know about this three months in advance—as soon as they receive a dismissal note. As a result, the first meeting with a caseworker can take place before the actual unemployment entry as well. (In our RCT, however,

⁵ The law covering the policy throughout the 2010s is written in the Social Code II Section 15 and the Social Code III Section 37.

all caseworkers were instructed to conclude the first IA only after the actual unemployment entry; see Section 3.)

The assignment of caseworkers to clients is quasi random and is typically governed by the first letter of the last name of the client, by the first or last digits of various codes that the individual bears, and by who is the first available caseworker when the client enters the agency for the first meeting. As a rule, the client keeps the same caseworker throughout his/her UI spell. There is no space for discretionary behavior by the caseworker regarding the contents of the IA. However, it is possible that the individual impact of the IA depends on the caseworker's behavior. We return to this in the results section. After the first meeting, the caseworker only updates the IA if strictly necessary, for example, if the unemployed hands in a disability note. Apart from this, the IA is typically updated after at least six months (for those aged 25+), to take changing circumstances and completed ALMP participation into account.⁶

If the UI recipient is found not to comply with the obligations and guidelines on search behavior and ALMP participation, whether they are mentioned in detail in the IA or not, then (s)he may receive a punishment in the form of a benefits reduction (i.e., a sanction). These are relatively severe, typically involving a full benefits withdrawal for at least one week, where the length of the period depends on the type of violation. A second detected violation may lead to a complete UI benefits withdrawal for more weeks.

Various countries have adopted policies that share features with the IA policy. In 2016, the Council of the EU issued guidelines to the member states on the policy mix to integrate the long-term unemployed into the labor market (EU, 2016). One aspect of this concerned "job integration agreements," which should define goals and obligations of the unemployed person and the counseling, help, and support measures by the service provider. A recent critical interim assessment (EU, 2021) observes that countries have taken varied approaches and complains that many countries are actually slow to implement it and many do not have a full-blown system. Often, the "job integration agreements" in place only prescribe counseling, assistance, training, and the provision of information to the unemployed, sometimes complemented with advice on childcare services, health services, and debt counseling (see also European Commission, 2019). The German IA policy is praised as a policy that fulfills the original EU guidelines. In this sense, our evaluation provides valuable insights for the other EU countries.

2.3. Self-Reported Perceptions of IA among Caseworkers. To gauge caseworkers' perceptions and assessments of the IA for UI recipients, we held a short survey in June 2012, that is, one month before the RCT began, among caseworkers in the participating agencies.⁷ Unit nonresponse was 28% and there was also substantial item nonresponse, resulting in a total of 159 respondents who answered each question used in this subsection.

The survey was set up as a list of statements for each of which the caseworker could indicate his or her agreement. We observe that 16% of the respondents agree mostly or fully with the statement that IAs are supportive for the job seekers in their search for work. Next, 19% agree mostly or fully with the statement that IAs help the job seekers to claim their rights. Conversely, 74% state that they use the IA at least to some extent to control the effort by the job seeker (i.e., to monitor the job seeker). These numbers confirm the descriptions of the nature of the IA in Section 1 and Subsection 2.2. Regarding the contents of the IA agreement, the caseworkers' responses support our above descriptions as well.

The survey also reveals that caseworkers envisage IA effect heterogeneity. On average, they believe that IAs do not increase the reemployment probability of individuals who have a good connection to the labor market and who can be expected to find work on their own within half

⁶ Before or after the IA is signed, caseworkers may profile their clients according to their assessment of the support they need. These are not in the data used in this article but our extended analysis allows effects to differ by an index of individual characteristics and labor market history.

⁷ Preliminary findings from this survey were reported in German in van den Berg et al. (2014).

a year. They tend to view IAs as more useful for individuals who in their view need to be activated and/or receive job search assistance or training. Of course, it is not clear whether the views on the usefulness of such support precede the views on the usefulness of an IA. But it appears that the usefulness of an IA is regarded to be higher if the individual does not have excellent prospects.

We also conducted a short survey among a sample of participants in the experiment. Again, the unit and item nonresponse was sizeable. More importantly, unit nonresponse was not balanced across treatment groups,⁸ which is why we mostly do not use the survey responses. The survey was again set up as a list of statements for each of which the respondent could indicate his or her agreement. The survey was held around 1.5 months after unemployment entry and the questions relating to the IA were only put forward to respondents who (and were assigned to have) received the IA in the first month. In the resulting small subsample of 127 individuals, less than half (44%) agree mostly or fully that IAs are supportive in their search for work. However, a much larger fraction (80%) feels that the IA serves as a reminder of their obligations during their search for work. And 78% agrees with the statement that the IA is a tool with which the caseworker can control the individual (i.e., to monitor the job seeker). Here, it should be kept in mind that the respondents are informed that the survey is carried out by the Institute for Employment Research (IAB, which is the main research and data institute of the FEA) among employment agency clients. Although they are also informed that responses are strictly confidential, some may have given answers that they deem to be desired by the FEA, so that the actual assessment of IAs may be even more tilted toward monitoring and away from counseling.

3. EXPERIMENTAL DESIGN

3.1. *Treatment Arms.* We randomize two aspects of the policy: the timing of the IA and the advance notification of the IA. Randomization takes place at the individual level upon entry into unemployment. We allow for four treatment arms. In treatment arm A, the IA is supposed to be signed in the first month of unemployment. In treatment arms B and C, this is supposed to occur three months after entry (if the individual is still unemployed), and in treatment arm D the signing is supposed to take place for the first time six months after entry (again conditional on unemployment). Treatment arms C and D do not include an advance notification of the future IA. In contrast, treatment arm B involves the receipt of a written announcement during the first meeting with the caseworker, informing the individual about the requirement to sign an IA in the third month of unemployment. This includes a detailed description of the typical content of IAs. In addition to that, it states that noncompliance with the content of the IA may lead to a sanction in the form of benefits cuts (see Appendix 3 for the exact wording of the announcement). In other words, treatment arm B provides an advance announcement that the rules the individual will have to comply with will be announced and fixed within an IA in around three months.

Table 1 summarizes the treatment arms. Each of the four possible treatment statuses in the RCT is given a 25% assignment probability. The Social Code legal framework does not allow for an RCT with a treatment arm in which the individual is never confronted with an IA. Similarly, it was not possible to randomize parts of the contents of the IA, so we could not introduce random variation, for example, in the number applications per week or in the highest commuting time deemed acceptable for offers provided to the individual. Note, however, that this would have increased the number of treatment arms considerably, which would be impractical and would lead to underpowered inference at given sample sizes. However,

⁸ Replication of the estimation of treatment effects based on the subset of the administrative data that exclusively contains the potentially selective sample of survey respondents leads to results that are qualitatively different from those based on the full sample. Several of the estimates (but not all) are outside the confidence intervals of the estimates based on the main sample.

TABLE 1
EXPERIMENTAL DESIGN

Group	IA in Month	IA Announced
A	1	No
B	3	Yes
C	3	No
D	6	No

NOTE: IA: integration agreement. IA announced: written announcement on IA handed out in the first month of unemployment.

these policy variations represent compelling candidates for future field experiments that could enhance our understanding of how different components of the IA impact employment outcomes.

3.2. Implementation of the RCT. We set up the experiment in five regional employment agencies out of a total of around 180 nationwide.⁹ The agencies were selected on the following criteria. First, during the time of the experiment (2012–13), they hosted no other pilot projects, for example, for the evaluation of other active labor market policies. Second, during this time, they did not face any other organizational changes, restructurings, or mergers. Third, the regions they served should not be too small in terms of population, to safeguard the sample size. In June 2012, around 2.8% of all unemployed individuals in Germany were registered at one of the five agencies. Fourth, they had to be dispersed across East and West Germany and across rural and urban regions, jointly creating some representativeness. The unemployment rate averaged across the five agencies does not differ from the national average (6.7% vs. 6.8%; both measured in June 2012). However, unemployment rates range from 2.5% in a Bavarian agency to 12.0% in an East German agency in the RCT. The agencies were informed by the FEA that they were selected to participate in the RCT. To prevent that the agencies' performance ratings would be affected by the work for the RCT or by the outcomes of clients involved in the RCT, it was communicated that RCT participation would not affect their performance goals.¹⁰

At each of the five agencies, two representatives of the FEA and of the research team presented the RCT to the agency head. FEA experts conducted instruction lessons with team leaders of caseworker teams in participating agencies before the project started (teams usually consist of 5 up to 15 caseworkers). The caseworker team leaders, in turn, instructed single caseworkers. The research team designed instruction material consisting of a presentation, a FAQ list, and a two-sided plastic slide summarizing the experimental design which was meant to be placed on each caseworker's desk throughout the experiment. The presentation highlighted the importance of the research question and why it could only be answered by means of an RCT. The material included verbal and graphical descriptions of the treatment arms. The target population was described and it was emphasized that other elements of the placement process were not supposed to differ across treatment groups, and in particular that all groups should have the same degree of access to ALMP instruments. Follow-up information was made available by e-mail and telephone.

The target population of the experiment is the full set of new entries into unemployment in the five employment agencies between July 2012 and January 2013. Individuals who were

⁹ The RCT design was approved after an internal review by the IAB Project Approval board and after a critical review by the legal department of the FEA, without the imposition of any additional constraints on the proposed design.

¹⁰ Each year, the FEA headquarters agrees on targets for the regional directorates through framework agreements with the Federal Ministry of Labor and Social Affairs. The regional directorates then agree on corresponding targets with their local employment agencies. These targets cover aspects such as the share of integrations into employment and the average duration of unemployment (Kaltenborn et al., 2010). If the local units do not meet targets, regional directorates may suggest measures for improvement (Kaltenborn et al., 2010).

eligible for UI were supposed to participate in the trial, where those aged below 25 or registered as unemployed anytime in the quarter prior to the current unemployment spell were excluded. This is because those categories faced different institutional environments and/or placement processes. We also exclude females because parental leave is not observable in the data and cannot be identified as distinct from unemployment. Parental leave spells can take up to three years and are usually taken up by the mother of the child instead of the father.¹¹

In the first meeting between the caseworker¹² and a newly unemployed individual in the target population, the latter was randomly assigned with equal 25% probabilities to one of the four treatment arms.¹³ The randomization is triggered by the caseworker during the meeting. The caseworker had to open an app and enter the client's identification number, name, and date of birth into a computer system. Both the app and the system were developed by the FEA for evaluation studies. The system generates a random number (not based on above characteristics) which then determines the assigned treatment status. In the RCT, the assigned status was immediately displayed in the caseworker app and the caseworker had to acknowledge it by entering it into the usual placement software program. This stores the time and the randomization outcome as well as anonymized identifiers of the client and the caseworker. Caseworkers were not able to manipulate the randomization, for example by rerunning the randomization. Importantly, the unemployed individuals were not informed about the RCT.¹⁴

It is also important to note that the protocol specified that the content of the IAs cannot be influenced by the randomized treatment. However, it is possible that the content of the IAs that were signed in later months (typically six months or later after inflow) systematically differs from the content of earlier IAs. We regard such potential differences as part of the treatment. It is also possible that the treatment assignment influenced the frequency of subsequent meetings between the unemployed and the caseworker or that it influenced ALMP access. We will address this aspect in more detail below.

4. DATA

4.1. Administrative Databases. The empirical analysis uses administrative data of the IAB of the FEA.¹⁵ These consist of individual records for the full labor force, notably from the so-called IEB. The IEB contains sociodemographic individual characteristics and detailed employment and unemployment histories, including daily earnings, transfer payments, and participation in ALMP, sanctions, and meetings with caseworkers. The IEB does not contain detailed information about working hours and self-employment but we observe self-

¹¹ Schönberg (2009) develops an algorithm for detecting maternal leave in the same IEB (integrated employment history register) data that we use as well, but this presupposes that the mother is in an employment relationship at the onset of the leave period, whereas in our setting the leave period would start during an unemployment spell. Note that even in the absence of these issues, the sample size for women would be substantially smaller than the size of our sample of men, causing any analysis of the former to be underpowered.

¹² In our experiment, all caseworkers were instructed to deal with the IA and IA-related issues only after the actual unemployment entry.

¹³ In an additional fifth group, the unemployed were assigned to be treated "as usual" with respect to the IA. Adding this group was a request of the involved department at FEA headquarters. The treatment "as usual" typically corresponds to an early IA during the first meeting with the caseworker. There were no clear instructions for the caseworkers for this group, so it is hard to interpret findings for this group, and, indeed, the outcomes for this group may be affected by the ongoing RCT. Therefore, we exclude this group from the analysis. However, we include a brief presentation of our findings for this group in a sensitivity analysis in Subsection 6.3.

¹⁴ In such evaluation designs at the FEA, the default would be to obtain informed consent. This can be disposed of if this would plausibly induce selection and if there is no convincing prior evidence that a participant will be worse off because of participation. If consent is not required, informing the participants about the experiment can be disposed of if the latter would plausibly induce changes in behavior and could thus invalidate the RCT.

¹⁵ We use databases called IEB version V12.01.00 and ASU-EEI version V06.09.00-201604. These are social data with administrative origin which are processed and kept by IAB according to Social Code III. The data contain sensitive information and therefore are subject to the confidentiality regulations of the German Social Code (Book I, Section 35, Paragraph 1).

employment subsidies and whether a job is full-time or part-time. The data also include a variable capturing the day at which an IA is signed, which is important to validate whether the caseworker follows the experimental protocol.

The IEB records are merged at the individual level with the variables that are recorded by the computer system used for the randomization. In our analysis, we use the assigned treatments recorded by the latter system. Recall that this also provides anonymized identifiers that enable the mutual linkage of unemployed sharing a caseworker.

The main outcome variable is the duration from the start of the unemployment spell to the beginning of the first subsequent employment spell. The start of the unemployment spell corresponds to the first day of UI receipt (or the first day of being registered as a job seeker without some parallel employment, if that day occurs before the randomization).¹⁶ The duration outcome as defined above might include intermittent periods in which an individual is not registered as unemployed and does not receive any benefits from the FEA. For expositional convenience, we nevertheless refer to this as part of the unemployment duration. We exclude one individual from the sample because randomization occurred on a day outside the experimental time window. We exclude seven individuals who could not be unambiguously matched to administrative records. This leaves us with an estimation sample of 4,163 entrants into unemployment, with groups A, B, C and D containing 1,061, 1,013, 1,068 and 1,021 individuals, respectively. Descriptive statistics are in the next subsection.

Across the five regional employment agencies, 213 caseworkers participated in the experiment. Some of these may have worked part-time. On average, each caseworker dealt with 20 RCT participants, where the number per caseworker ranged from one to 76.¹⁷ See Figure A.1 in the Appendix for the distribution of RCT participants across caseworkers.

We finish this subsection by listing data sources that we do not have access to but that might have been useful to study. First, we do not observe the content of the IAs at the individual level. Second, we do not observe whether the IA is unilaterally signed. Third, we do not observe caseworker characteristics beyond an anonymous identifier. Fourth, we do not observe this caseworker identifier for clients who do not participate in the RCT. These limitations are motivated by costs of digitization as well as by requirements to protect confidential information and privacy. Fifth, caseworkers virtually never gave consent to merge their survey data records with administrative records of their clients. The caseworker survey data could not be merged with other data sources either.

4.2. Balancing Tests and Timing of the IA. Since caseworkers could not manipulate the randomization tool, we do not expect substantive differences between the four treatment groups in their prerandomization characteristics. To proceed, we perform a range of separate regressions in which individual prerandomization characteristics are regressed on three binary indicators of the treatment statuses A, B, and C (leaving out D as the reference status). Judged on the basis of *F*-tests for the three coefficients, almost every characteristic is well-balanced across the four experimental groups. Table 2 presents the distribution of selected characteristics across the four groups and the corresponding *p*-values for the balancing tests. Table A.1 in the Appendix provides results for additional characteristics including labor mar-

¹⁶ For individuals who are not registered as unemployed or as job seeking on the day of randomization, we define the start of the unemployment spell to equal the day of randomization. Individuals who are still employed on the day of the randomization are excluded from the sample, as the entry into unemployment after randomization may be endogenous among them. We also exclude individuals who were unemployed for more than six weeks on the day of randomization because such a pattern is hard to reconcile with the guidelines on the timing of the first meeting and/or with the experimental protocol. This last step reduced the sample size from 4,609 to 4,163.

¹⁷ This caseload refers to the participants in the experiment to the extent that they are used in the analysis. It is possible that the caseworkers concurrently dealt with job seekers who did not participate in our experiment, for example, because they entered unemployment before or after the period of randomization or because they had a previous unemployment experience shortly before the current entry into unemployment or because they fell outside of the sampling criteria, for example, because of their gender. We cannot match such unemployed individuals to the 213 caseworkers in our experiment.

TABLE 2
BALANCING—DISTRIBUTION OF SELECTED OBSERVED CHARACTERISTICS ACROSS EXPERIMENTAL GROUPS

	Treatment group (<i>N</i> = 4163)				<i>p</i> -Value
	A	B	C	D	
Age	41.7	42.1	41.6	41.3	0.480
Vocational training	0.693	0.715	0.700	0.699	0.738
University degree	0.090	0.081	0.089	0.102	0.431
Abitur (High-school degree)	0.147	0.145	0.154	0.177	0.163
German	0.892	0.885	0.880	0.899	0.536
Turkish	0.032	0.034	0.035	0.035	0.979
Previous wage	66.278	66.592	65.776	66.065	0.966
Duration of previous employment spell	639.531	629.747	637.778	662.849	0.632
Duration of previous nonemployment spell	119.012	119.934	103.491	113.112	0.217
Share in unemployment last five years	0.139	0.141	0.135	0.133	0.573
Subsidized self-employment in the last five years	0.036	0.046	0.045	0.033	0.331
Subsidized employment in the last five years	0.123	0.116	0.127	0.125	0.887
ALMP in the last five years	0.416	0.396	0.404	0.390	0.656
Average wage in the last five years	63.887	64.424	64.325	63.023	0.802
Number of observations	1061	1013	1068	1021	

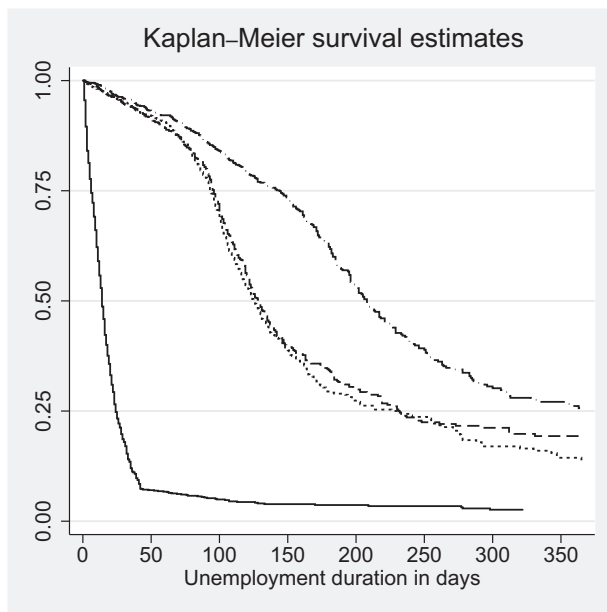
NOTE: Treatment A/C/D: integration agreement in month 1/3/6. Treatment B: integration agreement in month 3 with written announcement in month 1. *X* variables measured at the day of randomization. The previous wage corresponds to last observed daily wage in euro before the entry into unemployment. ALMP includes participation in training programs, public employment schemes, wage subsidies, mobility subsidies, and job search assistance programs. We report *p*-values associated to *F*-statistics from regressions of each variable on three treatment group dummies and a constant.

ket history indicators. Those results confirm that the randomization worked well. As a more encompassing way to examine the same issue, we estimate a multinomial logit model for the four treatment statuses as functions of the individual characteristics. This gives a *p*-value of 0.42 for the ensuing likelihood ratio test statistic of the null hypothesis of all coefficients of the characteristics being equal to zero, confirming that the observed characteristics are balanced across the different treatment arms. In such a context, Deaton and Cartwright (2018) argue that significance tests are mainly relevant if one is worried that the randomization has failed whereas otherwise it makes more sense to assess the magnitude of realized imbalances. In our case, it turns out that the differences in associations with observed characteristics are rather small.

In the RCT, the exact timing of the IA was not under our perfect control. In practice, the date at which the IA is signed depends on when meetings between caseworker and client are held, and the latter is subject to variation, for example, due to sickness absence and holidays. To assess this empirically, one may consider the estimation of Kaplan–Meier survival functions for the duration until the IA by the different treatment groups. Unfortunately, the interpretation of the estimates is problematic, as the durations until the IA are right-censored by exit to employment. One could assume independent right-censoring (conditionally on observed covariates) as an identifying assumption for the effect of the treatment status on the duration until IA. But this would mean that there are no unobserved confounders driving both the duration until IA and the duration until employment. Such an assumption is difficult to justify, as caseworkers may fine-tune the timing based on worker characteristics that are unobserved in the data.¹⁸

With this in mind, we merely provide some indicative statistics. In 25% of the cases where the IA takes place, the difference between intended and actual date exceeds one month. Figure 1 plots the Kaplan–Meier estimates. Clearly, they differ strongly across the treatment groups, which is of course to be expected. For example, after two months, less than 10% of

¹⁸ Note that if the assumption is valid then one can study the average effects of the timing of the IA with nonexperimental methods.



NOTES: Solid: IA in month 1 (Group A), long dash: IA in month 3 with announcement (Group B), dot: IA in month 3 without announcement (Group C), dash dot: IA in month 6 (Group D). Number of observations: 4,163. Please note that these are not consistent estimates of (one minus) the duration distributions until signing an IA because job exits create dependent right-censoring (see Subsection 4.2). Therefore, the interpretation should proceed with extreme caution.

FIGURE 1

KAPLAN-MEIER ESTIMATES OF SURVIVAL UNTIL SIGNING AN IA (IN DEPENDENT RIGHT-CENSORING SETTING)

group A is estimated to not have signed an IA yet. Furthermore, the estimated survival functions for the duration until IA are virtually identical across the groups C and D in the first two months of unemployment. If this were not the case then that would cast serious doubt on the implementation of the experimental design. The estimated functions for B and C are virtually identical throughout, which may be tentatively interpreted as confirming that those who anticipate the IA at three months do not use this knowledge to manipulate the timing of the IA.

Perhaps more surprisingly, the estimated functions show some variation within treatment groups, reflecting IA scheduling deviations. By exploiting the caseworker identifier variable, we find that such deviations are more common for some caseworkers than for others. Since the timing of meetings is primarily determined by the caseworker, this suggests that scheduling deviations primarily originate from the caseworker's views or attitude or from caseworker health shocks of varying degrees of severity and persistence (see, e.g., Schiprowski, 2020). As a sensitivity analysis, one might therefore drop all the clients of caseworkers with relatively many extreme deviations from the sample. Alternatively, one might estimate models allowing for interactions between the treatment status and an indicator of the caseworker's propensity to have scheduling deviations.

5. METHODOLOGICAL CONSIDERATIONS

5.1. Outcomes. The empirical analysis of the RCT faces a number of challenges that are common in the case of survival outcomes. First, note that ideally one would like to know effects on conditional reemployment rates at various elapsed durations t , as such rates are more closely related to behavior at t than, for instance, survival probabilities at t . However, with treatments affecting reemployment before any $t > 0$, randomization is lost if we condition on survival at some $t > 0$, as the composition in terms of unobserved characteristics will system-

atically differ across treatment arms (see, e.g., Abbring and van den Berg, 2005). Therefore, the comparison of reemployment rates in different treatment groups at some elapsed duration $t > 0$ does not allow for meaningful causal inference if the treatments may affect reemployment differentially before t .

This has a number of implications in our setting. It is conceivable that in groups A and B the treatments lead to group-specific behavior from the onset, so that the hazard rates in group A or B cannot be meaningfully compared to the hazard rates in any other group at any $t > 0$. In contrast, individual behavior should on average be identical across groups C and D until three months. Following insights from van den Berg et al. (2020), nonparametric causal inference on the difference between the reemployment rates in C and D is then possible for t exactly equal to three months. After three months, the treatment regimes differ between C and D, so that causal inference on reemployment rates is not possible anymore. Also, following van den Berg et al. (2020), the discontinuities in reemployment rates within group C at three months and within group D at six months enable identification of a causal effect of the IA on the reemployment rate at exactly those points in time, under the assumption that no other events take place at those points in time that lead to a discontinuity in the individual hazard rates. This approach does not allow for causal inference on reemployment rates at any other value of t . In practice, even these limited opportunities for causal inference on reemployment rates are not feasible, as they would require IA meetings in C and D to take place at exactly three and six months sharp, respectively. The empirical variation around those dates precludes such inference. Because of this, our primary outcomes of interest are the unconditional probabilities of leaving unemployment within certain durations t .

A second common challenge, by analogy to Ham and LaLonde (1996), is that inference on postunemployment outcomes is hampered for the reason that those are only observed if exit to work occurs before the end of the observation window. Whether this condition is satisfied depends on the treatment status and on unobservables, so, again, randomization is lost. Because of this, we interpret differences in accepted wages only as suggestive evidence. In addition, we do examine the effects on *earnings* obtained in t periods after inflow into unemployment. We consider two outcomes: (i) only labor earnings and (ii) total earnings consisting of labor earnings and UI benefits. Both earnings outcomes are observed for every individual.

5.2. Anticipation of Future IA Date. The comparison between treatment arms B and C enables us to evaluate whether advance notification of the timing of an IA at three months affects outcomes. To understand the results, we study a job search model of unemployed workers who are exposed to an event (IA) at a duration τ (three months). As a starting point, we assume that the event is unattractive from the point of view of the worker in the sense that it imposes constraints on her behavior, from τ onward.

In the spirit of Mortensen (1986), consider an unemployed individual who searches sequentially for a job. Given a particular search effort s , job offers arrive according to the rate $\lambda \cdot s$. Offers are random drawings from a wage offer distribution $F(w)$. Every time an offer arrives the decision has to be made whether to accept it or to reject it and search further. Once a job is accepted, it will be held forever at the same wage. During unemployment, a flow of benefits b is received and a flow of search costs $c(s)$ has to be paid. The individual maximizes the expected present value of income over an infinite horizon. For convenience, we take the partial equilibrium model to be stationary apart from the event at τ . That is, b , $c(\cdot)$, λ and F are assumed to be constant over time. Also, the model determinants are taken to satisfy the usual regularity assumptions.

Behavior at durations $t < \tau$ depends on how much is known about the IA. If the individual does not know about the treatment then her behavior up to τ can be captured by a reservation wage ϕ_0 and an optimal search effort s_0 that are constant over time. If the individual anticipates the event at τ then the model is genuinely nonstationary (van den Berg, 1990) and behavior up to τ can be captured by differential equations for the reservation wage $\phi(t)$ and

optimal effort $s(t)$, derived from the following asset flow equation for the expected present value of income $R(t)$ with ρ denoting the discount rate¹⁹:

$$(1) \quad \phi(t) = \rho R(t) = \max_{s(t)} \left[\frac{\phi'(t)}{\rho} + b - c(s(t)) + \frac{\lambda s(t)}{\rho} \int_{\phi(t)}^{\infty} (1 - F(w)) dw \right],$$

where $R(t)$ decreases until $t = \tau$ and thus $\phi(t)$ decreases as well whereas $s(t)$ increases until τ . Compared to the setting with no knowledge about the future event, $\phi(t) < \phi_0$ and $s(t) > s_0$. In a nutshell, individuals who anticipate the event aim to avoid the reduced attractiveness of the search environment after τ by being less selective with respect to job offers and by searching harder, before τ .

In obvious notation, the reemployment (or hazard) rates up to τ in cases C and B can be expressed as

$$\theta_0(t) = \lambda s_0(1 - F(\phi_0)), \quad \theta(t) = \lambda s(t)(1 - F(\phi(t))),$$

respectively. Clearly, $\theta(t)$ increases until τ . This implies that the reemployment rate on the interval $(0, \tau)$ is larger in B than in C and that the difference increases as t increases. This is the first main finding of this subsection. The ranking of B and C extends to the unconditional reemployment probability for any interval $(0, t)$ with $t < \tau$.

Regarding treatment B, it is not difficult to show that the above equations imply that

$$(2) \quad \frac{\phi''(t)}{\phi'(t)} = \rho + \theta(t),$$

so $\phi'(t)$ and $\phi''(t)$ are both negative, implying that $\phi(t)$ decreases at an increasing pace until $t = \tau$. Likewise, $s(t)$ increases at an increasing pace until $t = \tau$. By integrating (2) over the interval (t, τ) , we obtain that $\phi'(t)$ can be written as

$$(3) \quad \phi'(t) = [\phi'(\tau)] \cdot e^{-\rho(\tau-t)} \cdot \Pr(T > \tau | T > t),$$

where $\phi'(\tau)$ is the left-hand side derivative at τ . This equation provides insight into the determinants of the extent of anticipation of the event at τ at a fixed value of $t < \tau$. After all, if $\phi'(t)$ is much below zero then this means that the individual is strongly modifying her optimal strategy in response to the future event. Now consider the three terms on the right-hand side. The first term $\phi'(\tau)$ captures how severe the change in the search environment at τ is, so it is a measure of the relevance of the event.²⁰ For our purposes, the second and third terms are more interesting as they capture anticipation of a given severity of the event at τ . The second term captures that the future event is more important at t if the discount rate is low. The third term captures that the future event is more important at t if the individual is unlikely to escape unemployment before τ . This term equals $\exp(-\int_t^\tau \theta(u) du)$ so it only depends on the reemployment rate. Of course, this in itself depends on the path of ϕ . After all, Equation (2) is not a recursive expression. As a first-order approximation, the second and third terms can be represented by $\exp(-(\rho + \theta(t))(\tau - t))$.

This suggests that, for a given adverse event at τ , we can expect a large difference in outcomes between treatment arms B and C if ρ is small and if reemployment rates $\theta(t)$ are small.

¹⁹ This follows from van den Berg (1990), incorporating an optimally chosen search effort along the lines of van den Berg and van der Klaauw (2006).

²⁰ This can be seen most easily in the special case where search effort is fixed at say $s \equiv 1$ and the event at τ is an increase of the job offer arrival rate from say λ_L to λ_R whereas nothing else changes after that. Then, from Equation (1), we have, coming from $t \uparrow \tau$, that $\rho R(\tau) = \phi'(\tau)/\rho + b - c(1) + \lambda_L \int_{\phi(\tau)}^{\infty} (1 - F(w)) dw/\rho$ and, coming from $t \downarrow \tau$, that $\rho R(\tau) = b - c(1) + \lambda_R \int_{\phi(\tau)}^{\infty} (1 - F(w)) dw/\rho$. This gives $\phi'(\tau) = (\lambda_L - \lambda_R) \cdot \int_{\phi(\tau)}^{\infty} (1 - F(w)) dw$ which is the change in the arrival rate λ times a measure of its relevance in the expected present value.

In practice, reemployment rates are an order of magnitude larger than commonly assumed values of the discount rate (e.g., average reemployment rates are around 2 per year whereas a commonly chosen value of ρ is 0.05 per year). This means that the individual employability (or, similarly, the probability to become long-term unemployed) is the key candidate for the study of heterogeneous treatment effects when comparing B and C on $[0, \tau)$. This is the second main finding of this subsection and it is based on a novel approach to the interpretation of nonstationary search models.

So far we have not modeled behavior after τ . At the individual level, that behavior is equal for B and C (and can be represented by ϕ_1 and s_1 that are constant over time). Therefore, the magnitude of the change in behavior at τ differs between B and C. With treatment arm B, the present value $R(t)$ is a continuous function at τ so the reservation wage does not change as time proceeds from just before τ (say, at $t = \tau^-$) to τ , so $\phi(\tau^-) = \phi_1$. With arm C, the event is unanticipated, so the perceived present value jumps downward at τ , and therefore the reservation wage jumps downward as well, from ϕ_0 to ϕ_1 . The latter leads to an upward jump (i.e., a discontinuity) in the reemployment rate at τ .

To use this for a test, we first need to address the fact that the reemployment rates are also affected by search effort. This in turn requires a more explicit discussion of the nature of the event at τ . In particular, the IA may be seen as imposing a minimum required search effort s^* which exceeds the value chosen in absence of the IA. In that case, the effort at τ will jump upward both in B and in C.²¹

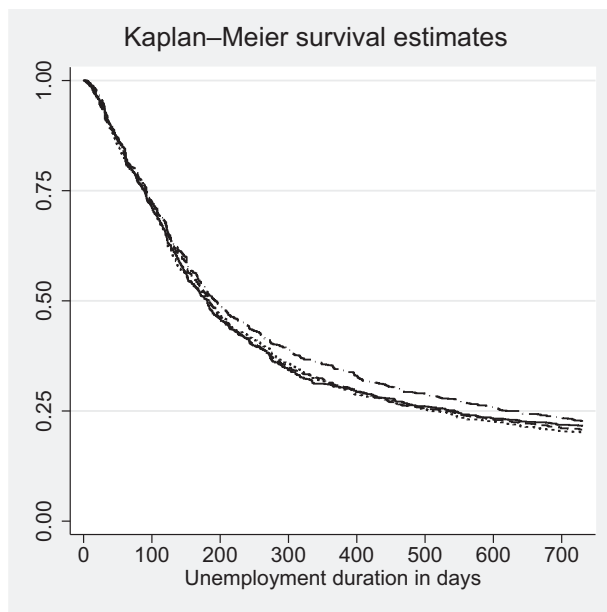
However, as we have derived that $s(t)$ exceeds s_0 at any $t < \tau$, the upward jump in effort is smaller in B than in C. Taking this together with the results on the reservation wage at τ , it follows intuitively that the upward jump in the reemployment rate for treatment arm C is larger than for treatment arm B. This is the third main finding of this subsection, and it leads to a test comparing the sizes of a discontinuity in the hazard rate at τ between groups B and C.

The first challenge for the implementation of this test idea is the issue discussed in Subsection 5.1. Dynamic selection due to unobserved heterogeneity may proceed at different speeds in groups B and C, precluding a clean comparison (quantitative causal inference) of hazard rates around τ . In an RCT, systematic unobserved characteristics at baseline are independent of the treatment status. It is not difficult to show that in that case, a discontinuity of the individual hazard rate at an elapsed duration τ is preserved under aggregation over unobserved heterogeneity. However, the ranking of the discontinuity sizes between groups B and C is not necessarily preserved as it depends on interactions between the treatment status and the unobserved characteristics in the individual hazard rates up to τ .

A second challenge is that, as discussed earlier, the timing of the IA is not uniform within treatment arms, so τ is dispersed within groups B and C. This complicates the inference based on hazard rates around τ . In particular, we do not observe the individual-specific value of τ if the individual leaves unemployment before τ . We therefore do not focus on the size of discontinuities but rather examine the steepness of the slopes of the empirical hazard rates around three months, and we treat findings based on shapes of the hazard rates around τ as tentative evidence only.

We finish this subsection with some more general remarks. First, as mentioned above, the IA event may include nudging elements leading to an increase of the job offer arrival rate and thereby an improvement of reemployment opportunities after τ . If individuals can acknowl-

²¹ The empirical evidence for a positive relationship between institutional requirements on search effort and the resulting reemployment rate is mixed. Van den Berg and van der Klaauw (2006) find a zero effect and show that this can be attributed to a substitution from informal into formal search channels. Johnson and Klepinger (1994) find a positive effect. Van den Berg and van der Klaauw (2019) argue that the size depends on labor market conditions. In addition, there might exist important displacement and equilibrium effects of such policies (Lise et al., 2004; Ferracci et al., 2014; Crépon et al., 2013). Toohey (2021) provides evidence that work search requirements increase search effort and reemployment at the individual level but may only reduce unemployment rates in times where unemployment is low.



NOTES: Solid: IA in month 1 (Group A), long dash: IA in month 3 with announcement (Group B), dot: IA in month 3 without announcement (Group C), dash dot: IA in month 6 (Group D). Number of observations: 4,163.

FIGURE 2

KAPLAN–MEIER ESTIMATES OF THE SURVIVAL FUNCTION UNTIL EXIT TO WORK

edge this benefit of nudging in advance then, before τ , this would make the future event less unattractive in the eyes of individuals in group B. This could mitigate the size of the differences between the effects of B and C before τ . Second, individuals in C may expect the IA event to occur at some rate η , in which case the IA has a so-called ex ante effect. This also tends to mitigate the size of the differences between B and C. Third, note that up to τ , the groups C and D behave identically on average, so for the above purposes D may be added to C on that time interval.

6. RESULTS

6.1. Average Effects. Figure 2 shows Kaplan–Meier estimates of the survival function until exit to employment, that is, estimates of the probability of having found a job as a function of the time t since the start of the unemployment spell. Note that we do not censor observations if they leave registered unemployment without entering employment directly. Therefore, the estimated survival rate at a duration t simply equals the ratio of the number of individuals at risk (i.e., who have not found a job yet) divided by the size of the corresponding treatment group. We discuss standard errors of estimated effects in binary-outcome analyses below, so the discussion of the estimated functions is brief. The estimated functions for the four groups are virtually indistinguishable in the first 120 days after unemployment entry. This suggests that signing an IA very early has on average no short-term impact on the probability of getting a job. At higher durations (around the median of about 200 days) individuals assigned to group D have a lower probability of having entered employment. There seem to be no systematic differences between groups A, B, and C.

Next, we estimate linear probability models. In what follows, we take treatment arm D (not-previously announced IA at six months) to be the reference category. The outcome y_{it} is

TABLE 3
EXIT TO WORK WITHIN 90, 180, 270, AND 365 DAYS AFTER UNEMPLOYMENT ENTRY

Until Day:	90		180		270		365	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
A	0.007	(0.019)	0.022	(0.021)	0.027	(0.020)	0.041**	(0.019)
B	−0.010	(0.019)	0.001	(0.021)	0.021	(0.020)	0.033*	(0.020)
C	−0.003	(0.019)	0.022	(0.021)	0.024	(0.020)	0.047**	(0.019)
Mean D	0.254		0.474		0.589		0.650	

NOTE: Linear probability models. Dependent variable is one if an individual has found a job within 90/180/270/365 days after unemployment entry. Number of observations: 4163. Group A: IA in month 1. Group B: IA in month 3 with announcement at first meeting. Group C: IA in month 3 without announcement. Reference group is D: IA in month 6. Significance levels: *10%, **5%, ***1%. Individual controls included but not shown: age, nationality, education, previous wage, handicap, and previous employment history.

a binary indicator which is one iff an individual i moved to subsidized or unsubsidized work before t , and $A_i = 1$ iff i is assigned to group A, etc.

(4)
$$y_{it} = \beta_0 + A_i\delta_A + B_i\delta_B + C_i\delta_C + \varepsilon_{it}.$$

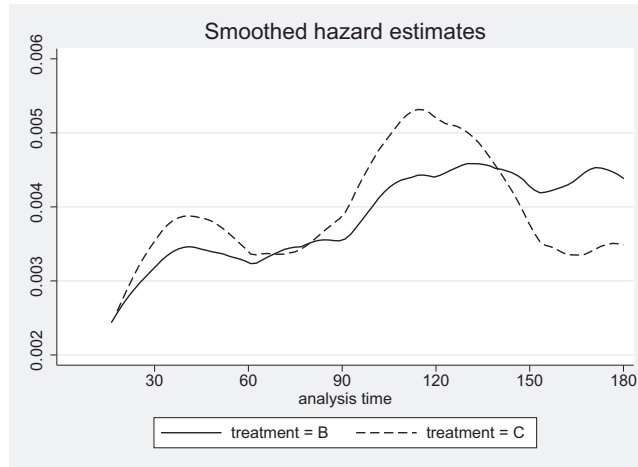
We also estimate versions including a vector x_i containing individual characteristics like age, nationality, education, last observed daily earnings, and other labor market history indicators. Table 3 reports the latter results, for t equal to 90, 180, 270, and 365 days. Not surprisingly, the results without x_i are virtually identical to those in the table.

The coefficients for A, B, and C are close to zero and insignificant at 90 days after entry into unemployment. At $t = 180$ and $t = 270$, the differences are not statistically significant either. The point estimates for effects at 270 days are around 2–3 percentage points for A, B, and C as compared to D. At one year, the effect estimates range from 3 to 5 percentage points; these are statistically significant at the 5% level for A and C and at the 10% level for B. Thus, on average, being assigned to a late IA reduces the probability of reemployment within a year by about 4 percentage points, from 69% to 65%, and it commensurately increases the probability of long-term unemployment. On average, it does not matter at any t whether the IA is signed immediately or after three months.

None of these results suggests that it matters much whether the IA at three months is announced in advance or not. To scrutinize this in more detail, we use information in the data on the exit rate to work around $t = 90$ for groups B and C. In line with the approach proposed in Subsection 5.2, we examine the steepness of the slopes of the empirical hazard rates around three months. Figure 3 displays kernel hazard estimates for B and C for durations up to six months.²²

This figure indicates that the hazard rate for B increases less steeply than for C, around 90 days, although the difference is not overwhelming. The result fits the theoretical prediction and thus supports the notion of anticipatory behavior. Individuals who are not informed in advance about the IA adjust their behavior more abruptly upon the signing of the IA, leading to a larger increase of the exit rate to work than among those who are informed in advance. However, this is not a quantitatively important phenomenon, as we do not find evidence of a larger reemployment probability for B at 90 days (or beyond) in Table 3.

²² The bandwidth is 14 days. In our data, the usual starting day of a new job is a Monday. For this reason, the bandwidth should be a multiple of seven days. Furthermore, in any given month, there are two moments where reemployment is much more frequent than at other moments: at the beginning of the month and in the middle of the month. Specifically, in our data, around 39% of the employment spells start in the first three days of a calendar month, and around 12% start at day 13, 14, or 15. This implies that a two-week bandwidth is more appropriate than a one-week bandwidth. A bandwidth of three or more weeks is too large in comparison to the 13 weeks that span the period from entry until the moment where the treatment groups B and C receive their IA. Indeed, according to graphical inspection, a bandwidth of three or more weeks leads to oversmoothing of the hazard rates around 13 weeks.



NOTES: Epanechnikov Kernel estimates for exit rate to work, with a bandwidth of 14 days. Solid: IA in month 3 with announcement (Group B), dashed: IA in month 3 without announcement (Group C). N : 2,081.

FIGURE 3

KERNEL HAZARD ESTIMATES FOR TREATMENT GROUPS B AND C

6.2. Heterogeneous Effects.

Employability For policy reasons, it is interesting to know if there are certain identifiable types of individuals whose reemployment benefits strongly or does not benefit at all from the timing and/or prior announcement of IAs. A key result from the theoretical analysis in Subsection 5.2 is that an individual's employability is the prime candidate for the study of heterogeneous treatment effects, in particular when comparing treatment arms B and C.²³ The caseworker survey (see Subsection 2.2) suggests that caseworkers often do not regard IAs as useful for the reemployment chances of individuals who are thought to find work on their own within half a year. In contrast, they see more potential for IAs in the case of individuals expected to need some help to bring them back to work. This also points at effect variation by employability.

We do not directly observe individual employability or caseworkers' expectations on employability in our sample. However, we may obtain an indicator of individual employability by predicting individual unemployment durations in terms of individual characteristics and labor market history. Instead of considering many possible employability types, we consider a binary classification. For this, we estimate a duration model on a different but similar sample. The estimated duration model is then used to classify individuals according to whether the predicted median duration exceeds six months or not.^{24,25}

²³ More generally, the behavior of individuals with low employability may be more restricted by the controlling aspects of IA.

²⁴ This approach can be seen as a profiling exercise. In practice, caseworkers themselves may profile their clients into one of a few categories. Such profiling is soft in the sense that it does not rely on an algorithm but on the caseworker's observation of the client's characteristics and history, the caseworker's subjective impressions, and the caseworker's assessment of the support that the client may need most. Clearly, the resulting category is related to whether the caseworker expects the unemployed individual to return to employment on his or her own within six months. At the macro level, about half of the inflow of unemployed is classified as being able to return on his or her own within six months. The data set we use for the present article does not contain comprehensive and directly usable information on profiling outcomes. Note that profiling classifications after the IA may have been made in response to the assigned treatment.

²⁵ A standard approach in the literature is to use the control group and to regress the outcome variable on a set of baseline characteristics and then to use this model to predict the potential outcomes for the full experimental sample. Based on that one can stratify the sample into groups with different levels of expected outcomes. Abadie et al. (2018)

Specifically, we estimate a descriptive Weibull proportional hazard model for the duration until employment given individual characteristics and labor market history x , so in obvious notation, $\theta(t|x) = \alpha t^{\alpha-1} \exp(x'\beta)$. The median $m(T|x)$ of T given x is then equal to

$$m(T|x) = (\log 2)^{\frac{1}{\alpha}} \exp\left(-\frac{x'\beta}{\alpha}\right).$$

It is not difficult to show that in this model, $\mathbb{E}(T|x) = m(T|x) \cdot \gamma$ for some $\gamma > 0$ that does not depend on x or β . Thus, a low median is equivalent to a low expected duration. Note that the individual predicted median duration $m(T|x)$ is a monotonic function of the single index $x'\beta$, so the binary outcome $I(m \geq 6)$ should give an employability classification that is relatively robust to misspecifications of the prediction model and to changes of the threshold value.

We estimate the prediction model with data we obtained covering all inflows into unemployment in the same regions in the year 2011, that is, from before the RCT. This is motivated by the fact that 2011 and 2012 are comparable years in terms of labor market conditions and in terms of absolute levels of flows into and out of UI among individuals aged 25-64 in the five regions (see Statistics of the FEA, 2019). Conditions in these two years 2011 and 2012 were slightly more favorable than in the surrounding years. Indeed, along the above dimensions, 2011 and 2012 are more similar to each other than to any of the surrounding years since 2009. The year 2011 was slightly more favorable than 2012, but the relevant flows differ only up to about 5% between the two years. Even when considering regions and 10-year age groups separately, differences in relevant flows between the two years do not exceed 10% in any subgroup.

The 2011 sample consists of 55,545 men aged 25–64. This is substantially larger than our RCT sample, because it covers a larger inflow window but also because the 2011 sampling design does not exclude some types of individuals or spells that would not be eligible for inclusion in the RCT, such as spells of individuals who had been unemployed at some point in the three months prior to the onset of the spell, or spells with meeting timing sequences deemed inadmissible for the RCT, or spells where individuals moved to work before the IA.²⁶ Table A.2 gives the estimation results for the prediction model.²⁷ Using the estimated prediction model, around 40% of our RCT sample are predicted to have a median duration below six months. (The next subsection contains sensitivity analyses regarding the six-month threshold value.) Table A.3 describes mean differences between covariates in the ensuing low- and high-employability groups in the RCT sample. The low-employability group with predicted medians above six months does not primarily consist of young unskilled workers but actually contains many older workers with higher previous wages and long previous employment spells, possibly with obsolete skills and coming from sectors in decline.

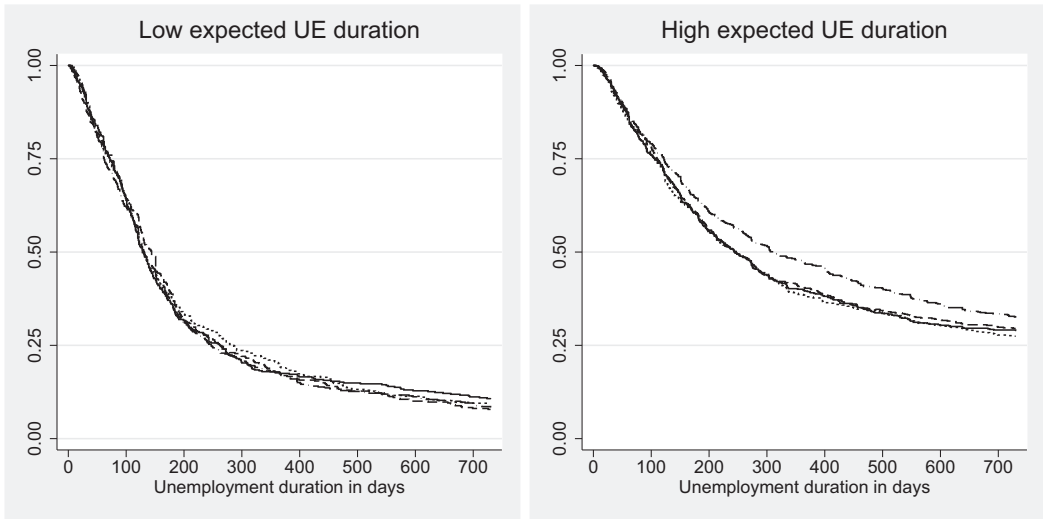
Some further comments are in order regarding the usage of the prediction model. First, spells that start in 2011 may be ongoing during the RCT, meaning that they may be affected by the execution of the RCT, even though the RCT is designed to avoid such externalities. More generally, it is undesirable if the prediction model is affected by outliers in the spell lengths. We investigate this empirically by artificially right-censoring spells at various points in time when estimating the prediction model.

Second, the predicted employability should relate to the views of the caseworkers in the RCT regarding employability because it reflects the experiences that the caseworkers accumulated before the RCT. The spells starting in 2011 were subject to the standard IA regime,

point out that this endogenous stratification can lead to substantial biases. Moreover, our sample sizes are modest, and, in fact, in our regions there is no natural control group during the RCT.

²⁶ Also, recall that the RCT sample excluded 20% of the inflow as they were randomized to not be in one of the four treatment arms.

²⁷ In the 2011 sample, we predict for 87% of the individuals who experienced a duration of more than six months that their predicted median is above six months. For 47% of those with a completed duration less than six months, it is predicted that the median is below six months. The latter result may reflect the restrictiveness of the Weibull model, which may lead to a slight underestimation of the prevalence of short durations.



NOTES: Individuals with predicted median unemployment duration ≤ 6 months on the left ($n = 1,688$) and individuals with predicted median unemployment duration > 6 months on the right ($n = 2,475$). Solid: IA in month 1 (Group A), long dash: IA in month 3 with announcement (Group B), dot: IA in month 3 without announcement (Group C), dash dot: IA in month 6 (Group D).

FIGURE 4

KAPLAN-MEIER ESTIMATES OF THE SURVIVAL FUNCTION UNTIL EXIT TO WORK—DEPENDING ON PREDICTED UNEMPLOYMENT-TO-EMPLOYMENT DURATION

meaning that the IA usually occurs during the first meeting. In a different regime (e.g., where everybody only receives an IA after six months), individual employability may change and caseworkers may respond to this. Whether such an equilibrium policy effect is quantitatively important depends on whether the regime affects the ranking of individuals in terms of their employability index.

Results by employability Figure 4 shows Kaplan-Meier estimates of the survival functions until exit to work. Among those with high employability, the estimated functions for the four treatment groups are very close. This suggests that, among them, signing an IA very early has on average no impact on their probability of getting a job. In contrast, among those with low employability, the survival function for group D is markedly different from the functions for groups A, B, and C, where the latter three are virtually equal. In particular, beyond 150 days group D displays a lower probability of having entered employment.

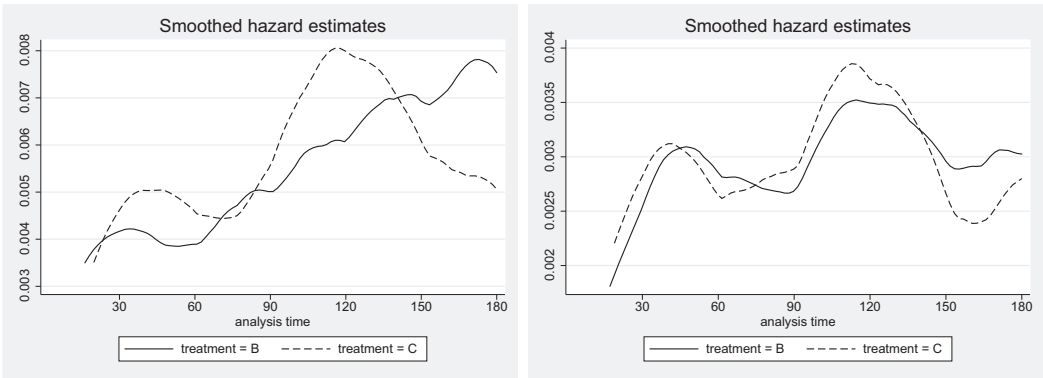
Table 4 presents estimation results for the regressions by employability. Among those with high employability, we do not find any statistically significant difference between treatment groups, regardless of the elapsed duration. The coefficients in the table have a negative sign, meaning that the probability of returning to work within a certain amount of time is highest for group D. Thus, early IAs are obviously not an effective tool to speed up reemployment for individuals with good labor market prospects. The same applies to the early notification of IAs.

This is different for those with lower employability. Here, early IAs in the first or third month of unemployment have statistically significant positive and quantitatively relevant effects on reemployment within nine months, as compared to having a later IA. For treatment groups A and C, the difference with D is even statistically significant at an elapsed duration as low as six months. One year after entry into unemployment, the differences between A, B, and C on the one hand and D on the other hand range from 6 to 9 percentage points; these differences are all statistically significant. Thus, on average, among those with low employability, being assigned to a late IA reduces the probability of reemployment within a year by

TABLE 4
EXIT TO WORK WITHIN 90, 180, 270, AND 365 DAYS AFTER UNEMPLOYMENT ENTRY—HETEROGENEOUS EFFECTS DEPENDING ON
PREDICTED UNEMPLOYMENT DURATION

Until Day:	90		180		270		365	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
<i>Predicted Median Unemployment Duration ≤ 6 Months^a</i>								
A	−0.024	(0.032)	−0.011	(0.032)	−0.021	(0.029)	−0.006	(0.026)
B	−0.027	(0.033)	−0.045	(0.033)	−0.025	(0.029)	−0.002	(0.026)
C	−0.032	(0.032)	−0.021	(0.033)	−0.040	(0.029)	−0.020	(0.027)
Mean D	0.344		0.642		0.771		0.823	
<i>Predicted Median Unemployment Duration > 6 months^b</i>								
A	0.030	(0.023)	0.046*	(0.027)	0.061**	(0.028)	0.075***	(0.027)
B	0.006	(0.023)	0.036	(0.027)	0.054*	(0.028)	0.060**	(0.027)
C	0.019	(0.022)	0.049*	(0.027)	0.064**	(0.027)	0.091***	(0.027)
Mean D	0.191		0.357		0.462		0.530	

NOTE: Linear probability models. Dependent variable is one if an individual has found a job within 90/180/270/365 days after unemployment entry. Predicted median unemployment duration is based on the coefficients of a hazard rate model estimated on an inflow sample into unemployment in the year before the experiment. Number of observations: ^a*N*: 1,688, ^b*N*: 2,475. Group A: IA in month 1. Group B: IA in month 3 with announcement at first meeting. Group C: IA in month 3 without announcement. Reference group is D: IA in month 6. Significance levels: *10%, **5%, ***1%. Individual controls included but not shown: age, nationality, education, previous wage, handicap, and previous employment history.



NOTES: Epanechnikov Kernel estimates for exit rate to work, with a bandwidth of 14 days. Solid: IA in month 3 with announcement (Group B), dashed: IA in month 3 without announcement (Group C). Left panel: sample of individuals with predicted median unemployment duration ≤ 6 months (*N* = 838). Right panel: individuals with predicted median unemployment duration > 6 months (*N* = 1,243).

FIGURE 5
KERNEL HAZARD ESTIMATES FOR TREATMENT GROUPS B AND C BY DEGREE OF EMPLOYABILITY

about 8 percentage points, from 61% to 53%. This is a substantial effect. For this, it does not matter whether an IA is signed immediately or after three months. The results also indicate that it is not quantitatively relevant whether the IA is announced in advance. To examine this in more depth, Figure 5 presents the equivalent of Figure 3 for each of the two employability groups. Each panel in the figure displays kernel hazard estimates for B and C for durations up to six months (bandwidths equal 14 days). Note that the vertical axis of the left panel (high employability) is more compressed than the vertical axis of the right panel.²⁸ Among high-employability individuals, we find a marked difference in

²⁸ Also note that the horizontal axis only covers the early stages of the spells. After 180 days the hazards decrease substantially.

the increase of the hazard rate around 90 days, with C displaying a much stronger increase than B (this is robust with respect to the bandwidth choice). According to the theoretical analysis, this means that individuals with treatment status B anticipate the IA at three months and modify their behavior before the IA in response to that. However, as in the full sample, this does not lead to a difference between B and C in the unconditional average reemployment probabilities at 90 days, so the anticipation is quantitatively unimportant. Among low-employability individuals, the hazard rates around 90 days are remarkably similar for B and C, and this does not provide much evidence of anticipatory behavior.

Tables A.4 and A.5 illustrate that the results do not change much when we decrease the right-censoring time point applied to the prediction sample.

Caseworker identifier As an additional heterogeneity analysis we interact the treatment effects with the caseworker identifier. Such an investigation can only have a limited scope due to (i) the large number of caseworkers and (ii) the fact that, although caseworker assignment is arguably quasi random, we were not able to randomize it within our RCT. However, interaction effects may be informative on the presence of heterogeneity in the extent to which a caseworker is able to put the IA to good use. If such heterogeneity is indeed present then this provides an incentive for the employment agency to let less-effective caseworkers learn from more-effective caseworkers.

In the data used for the above results, there are 13 caseworkers with each over 50 clients in the RCT. We estimate regression models in which the treatment effects are interacted with 13 corresponding binary caseworker indicators and where these indicators are also included as additive regressors, using the same data. Clients of caseworkers who had 50 or less clients in the RCT are the baseline category in these regression models. Among low-employability clients, we find strong evidence of effect heterogeneity according to an F -test (p -value is 0.041).²⁹ The standard deviations across the estimated treatment effects correspond to 0.42 for treatment A, 0.40 for treatment B, and 0.36 for treatment C, again suggesting a substantial amount of effect heterogeneity. The results are robust with respect to small changes in the threshold value of 50 clients per caseworker. For the reasons mentioned above, we do not zoom in further on sources of effect heterogeneity by caseworker. However, the results motivate further research to identify whether caseworkers can increase the reemployment effect of early IAs among low-employability clients by adopting the work practice used by the caseworkers whose clients display the largest effects.

6.3. Additional Outcome Variables and Sensitivity Analyses.

Wage-related outcomes Recall from Subsection 5.1 that inference of average treatment effects on initial wages in accepted jobs is not possible due to right-censoring of unemployment spells at the end of the observation window. With this in mind, Figure A.2 compares kernel density estimates for the initial wage (per day) in accepted jobs, measured at various duration endpoints. This does not suggest any large or systematic differences across the treatment groups.³⁰ Stratification by employability is not informative, as the degree of right-censoring differs starkly between the two subsamples. Moreover, it is not clear how to interpret wage effects by an employability measure based not on wages but on durations. As mentioned in Sub-

²⁹ In the subsample of clients with high employability, we do not find evidence of effect heterogeneity (p -value 0.65). Note that in this subsample we do not find an effect in a homogeneous specification either.

³⁰ The survey that was held among a subsample of RCT participants about 1.5 months after entry (Subsection 2.3) includes a question about the lowest acceptable wage (i.e., the reservation wage) for those still unemployed. It is difficult to use this information. As discussed in Subsection 2.3, the sample of respondents is nonbalanced. Moreover, the reservation wage is a determinant of being unemployed at 1.5 months. With these caveats in mind, we find no evidence that the observed reservation wages early in the spell are systematically different across treatment groups (see Table A.6). This is consistent with the absence of differences in accepted wages.

TABLE 5
ACTIVE LABOR MARKET POLICY PARTICIPATION, INVITATIONS TO MEETINGS AND VACANCY REFERRALS

	Predicted Median Duration ≤ 6 Months ^a						Predicted Median Duration > 6 Months ^b					
<i>Until Day</i>	<i>90</i>		<i>180</i>		<i>365</i>		<i>90</i>		<i>180</i>		<i>365</i>	
<i>Participation in ALMP</i>												
A	−0.000	(0.011)	−0.009	(0.012)	−0.008	(0.012)	0.004	(0.011)	0.007	(0.011)	0.003	(0.010)
B	−0.005	(0.011)	−0.010	(0.012)	−0.009	(0.012)	0.001	(0.011)	0.002	(0.011)	0.002	(0.010)
C	0.018	(0.013)	0.009	(0.013)	0.007	(0.013)	−0.003	(0.010)	−0.003	(0.010)	0.000	(0.010)
Mean D	0.052		0.071		0.080		0.056		0.063		0.066	
<i>Invitations to Meeting</i>												
A	0.000	(0.001)	0.000	(0.001)	−0.000	(0.001)	0.000	(0.001)	−0.000	(0.001)	−0.000	(0.001)
B	−0.000	(0.001)	−0.000	(0.001)	−0.000	(0.001)	0.000	(0.001)	0.000	(0.001)	0.000	(0.001)
C	0.001	(0.002)	0.001	(0.002)	0.001	(0.002)	0.001	(0.001)	0.000	(0.001)	0.000	(0.001)
Mean D	0.013		0.013		0.013		0.011		0.011		0.011	
<i>Vacancy Referrals</i>												
A	0.003	(0.005)	0.003	(0.005)	0.003	(0.005)	−0.003	(0.003)	−0.003	(0.003)	−0.003	(0.003)
B	0.007	(0.005)	0.006	(0.005)	0.006	(0.005)	−0.002	(0.004)	−0.001	(0.004)	−0.001	(0.004)
C	−0.004	(0.005)	−0.004	(0.004)	−0.005	(0.004)	−0.002	(0.003)	−0.001	(0.003)	−0.000	(0.003)
Mean D	0.054		0.050		0.050		0.036		0.032		0.031	

NOTE: OLS regressions. Outcome variables: sum of days in ALMP participation/sum of invitations/sum of vacancy referrals, each divided by the number of days spent in unemployment, by employability indicator. In the latter, the predicted median unemployment duration is based on the coefficients of a hazard rate model estimated on an inflow sample into unemployment in the year before the experiment. Number of observations: ^aN: 1,688, ^bN: 2,475. Group A: IA in month 1. Group B: IA in month 3 with announcement at first meeting. Group C: IA in month 3 without announcement. Reference group: IA in month 6. Significance levels: *10%, **5%, ***1%. Individual controls included but not shown: age, nationality, education, previous wage, handicap, and previous employment history.

section 6.2, the preunemployment wage is negatively correlated to the employability measure in the RCT sample.

Next, we consider effects on cumulative earnings. The average effects are reported in Table A.7. It turns out that we do not find any statistically significant impact on cumulative labor earnings or on the sum of cumulative labor earnings and UI benefits, measured at 180 or 365 days after entry into unemployment. Turning to heterogeneous treatment effects, our point estimates are more positive for unemployed with a rather long expected unemployment duration (see Table A.8). Whereas the estimates for treatment arms A and B are not statistically significant, the effect of receiving an IA in month 3 without previous announcement is statistically significant suggesting that cumulative labor earnings after a year are increased by around €1,240. When we consider the cumulative sum of labor earnings and UI benefits, the estimated effect of treatment C drops to around €926 after a year, which is still sizeable but is not statistically significant anymore. Overall, the effects are estimated with relatively low precision.

Usage of ALMPs According to the experimental protocol, caseworkers should not allow the assigned IA treatment to affect the frequency of meetings with the unemployed or their access to ALMPs. To verify this, we examine whether these are associated with each other, using the detailed information on meetings and ALMP participation in the data. Such analyses are descriptive as the observation of meetings and ALMP participation is restricted by the realized duration outcome. We regress the number of days spent in ALMP and the number of invitations divided by the days spent in unemployment on indicators for being assigned to treatment A, B, or C, controlling for observed background characteristics x . Analogously, we investigate whether the treatment groups differ in the probability of receiving vacancy referrals from the employment agency. Table 5 contains results by employability. None of the coef-

ficients is significantly different from zero. This suggests that our main findings are not driven by differences in the access to ALMP, the receipt of vacancy referrals or the number of meetings with the caseworkers.³¹

Moreover, we obtain qualitatively similar results when only considering transitions into unsubsidized jobs as transitions into employment, or when considering subsidized self-employment as part of regular employment (see Tables A.9 and A.10). We also investigate whether the probability of a recall is affected by the treatment. It turns out that recall differences across treatment groups are small and typically statistically nonsignificant (see Table A.11).

What the results in this subsection suggest is that IAs do not work by way of participation in other ALMPs. Also, the usage of other ALMPs does not seem to depend on the timing or advance notification of the IA. IAs thus appear to operate independently of other policy measures. IA effects can therefore be seen as policy effects that are separate from any effects of other ALMPs.

Treatment “as usual” Whereas our original research design consists of the comparison of four different treatment arms, the involved department of the FEA headquarters has implemented a fifth arm described as “the usual IA regime.” This should correspond to signing an IA early in the unemployment spell. In fact, there were some degrees of freedom in the organization of this fifth arm and there were no unified instructions across the involved local agencies. Therefore, we do not include this treatment arm in our main analysis. Tables A.12 and A.13 report results including this additional treatment “E.” We find, on average, a positive but statistically insignificant effect on the probability of leaving unemployment for a job. When we consider heterogeneous treatment effects, we find that the “treatment as usual” tends to have negative but statistically insignificant effects for those unemployed with a short expected unemployment duration, whereas it has a positive and statistically significant effect for low-employability individuals. Overall, the effects do not contradict the main results.

Alternative (sub)sampling criteria Next, we take a predicted median duration of seven months instead of six months as threshold for splitting the sample into two groups. The results are robust with respect to this (Table A.14). We additionally examine whether the results change when omitting individuals with a predicted median duration of more than three years until employment. It turns out that most point estimates are close to those for our main specification (see Table A.15). Finally, as a sensitivity analysis, we omit all clients of caseworkers with relatively many extreme scheduling deviations.³² The results are qualitatively the same (see Table A.16).

7. CONCLUSIONS

Signing an IA in the first or third month of unemployment (as opposed to later, in the sixth month) has on average a positive effect on entering employment within a year. More specifically, a late IA reduces the average probability of reemployment within a year by about 4 percentage points, from 69% to 65%, and it commensurately increases the probability of long-term unemployment. For this, it does not matter whether the IA is signed immediately or after three months.

³¹ In Table 5, we focus on the number of days spent in any ALMP. The overall share of days spent in ALMP is rather low (around 7%). The conclusions obtained for the aggregate measure do not change when we instead focus on the largest ALMP, which is the default training program.

³² For this purpose, we define that a caseworker *deviates* from the schedule if (i) a client in treatment group A is unemployed for 60 days or more and does not sign an IA before 60 days of unemployment, (ii) a client in treatment group B or C signs an IA before day 60 or after 135 days of unemployment, or (iii) a client in treatment group D signs an IA before day 135 or after day 245. In the sensitivity analysis, we exclude caseworkers who deviate in more than 40% of their cases.

A theoretical analysis based on job search models suggests that an individual's employability is the prime candidate for the study of effect heterogeneity, and this is corroborated by caseworker survey responses. It turns out that among those with high employability, the timing of the IA does not affect the probability of returning to work within any amount of time. If anything, early IAs have a negative effect on exits to work. Thus, early IAs are not an effective tool to speed up reemployment for individuals with good labor market prospects. This is different for those with lower employability. Here, early IAs in the first or third month of unemployment have significantly positive and quantitatively relevant effects on reemployment within 9 months and within 12 months, as compared to having a later IA. The differences are sometimes even statistically significant at an elapsed duration as low as six months. On average, among those with low employability, being assigned to an early IA increases the probability of reemployment within a year by about 8 percentage points, from 53% to 61% (so the relative increase is 15%). This is a substantial effect. For this, it does not matter whether an IA is signed immediately or after three months. Note that the positive overall effects of early IAs are exclusively driven by the lower-employability group.

We conclude from this that the IA is a valuable policy tool, especially for newly unemployed individuals with adverse labor market prospects, as it strongly reduces their probability of long-term unemployment. It is interesting to view this in the light of the existing evidence on the limited effectiveness of more traditional ALMP that, after all, often target those with adverse prospects. Our results suggest that the IA is an interesting new addition to the set of policies for this group. Apparently, an approach in which monitoring is presented in a rather neutral and formal fashion delivers desirable outcomes.

Our article contains a detailed theoretical analysis of anticipatory effects of advance announcements of the IA, and we develop an innovative econometric approach to detect such announcement effects. The corresponding empirical findings suggest anticipatory behavior in response to the advance announcement of an IA at three months. Upon the signing of the IA, individuals who are not informed in advance adjust their behavior more abruptly, as witnessed by a larger increase of their exit rate to work, than those who are informed in advance. However, this is not a quantitatively important phenomenon, as we do not find evidence of announcement effects on unconditional reemployment probabilities at any elapsed duration.

We also examine interaction effects with caseworker identifiers. Among low-employability clients, we find statistical evidence of effect heterogeneity. In our view, this motivates further research to identify whether early IAs for low-employability clients can be put to better use by adopting work practices used by the most effective caseworkers.

According to the experimental protocol, caseworkers should not allow the assigned IA treatment to affect the frequency of meetings with the unemployed or their access to ALMPs. We verified that this was indeed the case (and this also applies to the frequency of vacancy referrals), so that effects cannot be attributed to differential usage of other policy instruments. Also, results are robust with respect to the usage of wage subsidies or self-employment subsidies. All in all, IA effects appear to operate independently of other policy measures. We do not find effects on recalls or on accepted wages, where it should be kept in mind that the latter are only observed for uncensored unemployment spells.

First findings from our study were presented to the governing board of the German FEA. This led the FEA to implement a major modification of the usage of IAs in the UI system. Caseworkers conduct a soft profiling for individuals entering unemployment, distinguishing between unemployed persons who are considered to be able to find work by themselves within six months and those who are not. Job seekers profiled as part of the former group are now not exposed anymore to a mandatory IA in the first three months of unemployment. In the absence of aggregate numbers on the fraction of newly unemployed UI recipients with this subjective reemployment characteristic, we cannot quantify the number of individuals who directly benefit from this policy change. A crude indication could be based on the

annual inflow into UI (2.55 million in 2012; see Statistik der Bundesagentur für Arbeit, 2020) and the fraction with high employability in our data (41 %).³³

DATA AVAILABILITY STATEMENT While the data underlying our analysis are not completely proprietary, access to the data is restricted. The data we use are social data and partly of administrative origin; they contain sensitive information and are subject to confidentiality regulations. Getting access to the data through the research data center of the Institute for Employment Research (IAB) would require a contract with IAB. Of course, we would support researchers interested in replicating the results with the paperwork required to get data access.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information

REFERENCES

- ABADIE, A., M. M. CHINGOS, and M. R. WEST, “Endogenous Stratification in Randomized Experiments,” *Review of Economics and Statistics*, 100(4) (2018), 567–80.
- ABBRING, J. H., and G. J. VAN DEN BERG, “Social Experiments and Instrumental Variables with Duration Outcomes,” IFAU Working Paper 2005:11, Uppsala, 2005.
- ALTMANN, S., A. FALK, S. JÄGER, and F. ZIMMERMANN, “Learning about Job Search: A Field Experiment with Job Seekers in Germany,” *Journal of Public Economics*, 164 (2018), 33–49.
- BELOT, M., P. KIRCHER, and P. MULLER, “Providing Advice to Job Seekers at Low Cost: An Experimental Study on Online Advice,” *Review of Economic Studies*, 86(4) (2019), 1411–47.
- BLACK, D. A., J. A. SMITH, M. C. BERGER, and B. J. NOEL, “Is the Threat of Reemployment Services More Effective Than the Services Themselves? Evidence from Random Assignment in the UI System,” *American Economic Review*, 93 (2003), 1313–27.
- BLUNDELL, R., M. COSTA DIAS, C. MEGHIR, and J. VAN REENEN, “Evaluating the Employment Impact of a Mandatory Job Search Program,” *Journal of the European Economic Association*, 2 (2004), 569–606.
- BOOKMANN, B., C. OSIANDER, M. STOPS, and H. VERBEEK, “Effekte von Vermittlerhandeln und Vermittlerstrategien im SGB II und SGB III (Pilotstudie),” IAB-Forschungsbericht 7/2013, 2013.
- Bundesagentur für Arbeit, “Zugänge ins Job-to-Job Potenzial (100 Tage) untergliedert nach den festgelegten Integrationsprognosen,” Mimeo, Controlling Berichtswesen, 2017.
- BÜTTNER, T., “Ankündigungseffekt oder Maßnahmewirkung? Eine Evaluation von Trainingsmaßnahmen zur Überprüfung der Verfügbarkeit,” *Journal for Labour Market Research*, 41 (2008), 25–40.
- CARD, D., J. KLUVE, and A. WEBER, “What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations,” *Journal of the European Economic Association*, 16(3) (2018), 894–931.
- CRÉPON, B., E. DUFLO, M. GURGAND, R. RATHELOT, and P. ZAMORA, “Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment,” *Quarterly Journal of Economics*, 128(2) (2013), 531–80.
- , and G. J. VAN DEN BERG, “Active Labor Market Policies,” *Annual Review of Economics*, 8(1) (2016), 521–46.
- , M. FERRACCI, G. JOLIVET, and G. J. VAN DEN BERG, “Information Shocks and the Empirical Evaluation of Training Programs during Unemployment Spells,” *Journal of Applied Econometrics*, 33 (2018), 594–616.

³³ An alternative estimate may be constructed from the number of 0.6 million individuals who, in 2016, registered themselves as job searchers because they expected to lose their current job within 100 days and who were deemed to be able to find work by themselves within six months of unemployment (see Bundesagentur für Arbeit, 2017). However, it is not clear how informative this number is. On the one hand, it includes individuals who, in the end, remain employed. On the other hand, for an unknown number of individuals who registered themselves as job searchers because they expected to lose their current job within 100 days, the subjectively assessed employability is not recorded in these data while they were actually deemed to be able to find work by themselves within six months of unemployment.

- DEATON, A., and N. CARTWRIGHT, "Understanding and Misunderstanding Randomized Controlled Trials," *Social Science and Medicine*, 210 (2018), 2–21.
- DE GIORGI, G., "Long-Term Effects of a Mandatory Multistage Program: The New Deal for Young People in the UK", Working Paper, IFS, London, 2005.
- EBERWEIN, C., J. C. HAM, and R. J. LALONDE, "Alternative Methods of Estimating Program Effects in Event History Models," *Labour Economics*, 9 (2002), 249–78.
- European Commission, "Study Supporting the Evaluation of the Council Recommendation on the Integration of Long-Term Unemployed into the Labour Market," Directorate-General for Employment, Social Affairs and Inclusion, Brussels, 2019. <https://data.europa.eu/doi/10.2767/397387>
- EU, "Council Recommendation of 15 February 2016 on the Integration of the Long-Term Unemployed into the Labour Market," *Official Journal of the European Union*, C 67 (2016), 1–5.
- EU, "2020 Joint EMCO and SPC Review of the Council Recommendation on the Integration of the Long-Term Unemployed into the Labour Market," Council of the European Union, Brussels, 2021.
- FERRACCI, M., G. JOLIVET, and G. J. VAN DEN BERG, "Evidence of Treatment Spillovers within Markets," *Review of Economics and Statistics*, 96 (2014), 812–23.
- HAM, J. C., and R. J. LALONDE, "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, 64 (1996), 175–205.
- HECKMAN, J., R. LALONDE, and J. SMITH, "The Economics and Econometrics of Active Labor Market Programs," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Volume 3 (Amsterdam: North-Holland, 1999) 1865–2097.
- IMMERVOLL, H., and C. M. KNOTZ, "How Demanding are Activation Requirements for Jobseekers?" IZA Discussion Paper, 11704, 2018.
- JOHNSON, T. R., and D. H. KLEPINGER, "Experimental Evidence on Unemployment Insurance Work-Search Policies," *Journal of Human Resources*, 29 (1994), 695–717.
- KALTENBORN, B., N. WIELAGE, A. I. HENKEL, T. WEISHAUPT, H.-L. BUCHHOLZ, F. OSCHMIANSKY, and A. v. BOTHMER, "Zielsteuerung in der Arbeitsverwaltung - ein europäischer Vergleich," Forschungsbericht / Bundesministerium für Arbeit und Soziales, FB409, 2010. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-334678>.
- KNOTZ, C. M., "A Rising Workfare State? Unemployment Benefit Conditionality in 21 OECD Countries, 1980–2012," *Journal of International and Comparative Social Policy*, 34(2) (2018), 92–108.
- LACHOWSKA, M., M. MERAL, and S. A. WOODBURY, "The Effects of Eliminating the Work Search Requirement on Job Match Quality and Other Long-Term Employment Outcomes," Department of Labor, Washington, DC, 2015.
- LALIVE, R., J. VAN OURS, and J. ZWEIMÜLLER, "The Effect of Benefit Sanctions on the Duration of Unemployment," *Journal of the European Economic Association*, 3 (2005), 1386–417.
- LISE, J., S. SEITZ, and J. SMITH, "Equilibrium Policy Experiments and the Evaluation of Social Programs," NBER Working Paper 10283, 2004.
- MICHALOPOULOS, C., P. K. ROBINS, and D. CARD, "When Financial Work Incentives Pay for Themselves: Evidence from a Randomized Social Experiment for Welfare Recipients," *Journal of Public Economics*, 89 (2005), 5–29.
- MORTENSEN, D., "Job Search and Labor Market Analysis," in O. C. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, Volume 2 (Amsterdam: North-Holland, 1986), 849–919.
- O'FLYNN, J., "From New Public Management to Public Value: Paradigmatic Change and Managerial Implications," *Australian Journal of Public Administration*, 66(3) (2007), 353–66.
- SCHÖNBERG, U., "Does the IAB Employment Sample Reliably Identify Maternity Leave Taking? A Data Report," *Zeitschrift für Arbeitsmarktforschung* 42 (2009), 49–70.
- SCHÜTZ, H., P. KUPKA, S. KOCH, and B. KALTENBORN, "Eingliederungsvereinbarungen in der Praxis: Reformziele noch nicht erreicht", IAB—Kurzbericht, 18/2011, 2011.
- SCHIPROWSKI, A., "The Role of Caseworkers in Unemployment Insurance: Evidence from Unplanned Absences," *Journal of Labor Economics*, 38(4) (2020), 1189–225.
- Statistics of the FEA, "Entries into and Exits Out of Unemployment Benefit Receipt," Data Warehouse of the Statistics of the Federal Employment Agency, 2019, Accessed on October 9, 2019.
- Statistik der Bundesagentur für Arbeit, "Arbeitslosengeld (Zeitreihen) November 2020," Nuremberg, 2020.
- TOOHEY, D., "Work Search Requirements in General Equilibrium," Working Paper, 2021.
- VAN DEN BERG, G. J., "Nonstationarity in Job Search Theory," *Review of Economic Studies*, 57 (1990), 255–77.
- and B. VAN DER KLAUW, "Counseling and Monitoring of Unemployed Workers: Theory and Evidence from a Controlled Social Experiment," *International Economic Review*, 47 (2006), 895–936.
- , A. H. BERGEMANN, and M. CALIENDO, "The Effect of Active Labour Market Programs on Not-Yet Treated Unemployed Individuals," *Journal of the European Economic Association*, 7 (2009), 606–16.

- , B. HOFMANN, G. STEPHAN, and A. UHLENDORFF, “Was Vermittlungsfachkräfte von Eingliederungsvereinbarungen halten: Befragungsergebnisse aus einem Modellprojekt,” IAB-Forschungsbericht, 11/2014, 2014.
- , and B. VAN DER KLAUW, “Structural Empirical Evaluation of Job Search Monitoring,” *International Economic Review* 60 (2019), 879–903.
- , A. BOZIO, and M. COSTA DIAS, “Policy Discontinuity and Duration Outcomes,” *Quantitative Economics*, 11 (2020), 871–916.