Which Active Labor Market Policies Work for Male Refugees? Evidence from Germany

ZEIN KASRIN ⓘ **AND STEFAN TÜBBICKE** ⓘ

Institute for Employment Research (IAB), Basic Income Support and Activation (GSA), Regensburger Str. 104, 90478 Nuremberg, Germany
email: Zein.Kasrin@iab.de
email: Stefan.Tuebbicke@iab.de
Corresponding author, email: Zein.Kasrin@iab.de

Abstract

In this paper, we estimate the causal effects of a set of active labor market programs for male unemployed refugees on welfare who entered Germany between, 2013 and September, 2016. Using rich administrative data, we employ covariate balancing propensity scores combined with inverse probability weighting to estimate effects up to 33 months after the start of treatment. Our results show that relatively short-term training in the form of Schemes by Providers and In-Firm Training, as well as longer-term Further Vocational Training programs have a positive impact on both the employment chances as well as labor market earnings of refugees in the medium run. So-called "One Euro Jobs", a public employment program, does not yield positive effects on employment or earnings. Sensitivity analyses confirm that our results are unlikely to be driven by unobserved confounding.

Keywords: Refugees; active labor market programs; evaluation; inverse probability weighting

1. Introduction

According to official statistics from the German Federal Office for Migration and Refugees (2021), about 2.3 million persons applied for asylum between 2013 and 2020. Around 52 percent were submitted between 2015 and 2016 alone, mostly by asylum seekers from Syria, Afghanistan and Iraq. This recent refugee wave was caused by intense instability as a result of war, political and/or socioeconomic volatility, factors that are unlikely to be resolved in the foreseen future. Therefore, the integration of refugees in the German labor market is a major policy concern.

In addition to the importance of employment for economic purposes, studies have shown that it is vital for a person's social integration and well-being (Clark, 2003; Paul and Moser, 2009; Pohlan, 2019). For refugees, this effect is likely stronger given they faced displacement and need to integrate in a foreign

environment. Moreover, successfully integrating refugees into the German labor market could help in reducing labor supply shortages (Klinger and Fuchs, 2020). Recent evidence by Kosyakova et al. (2021) shows that almost 50 percent of refugees who migrated to Germany between 2013 and 2016 are in employment five years after arrival. Nevertheless, a large number of refugees remain unemployed, most of which receive welfare benefits (German Federal Employment Agency, 2021). Moreover, Brücker et al. (2020) shows that wages of employed refugees are often low, highlighting the poor quality of such employment.

Participation in active labor market programs has a significant potential in facilitating the employment integration of refugees. There is ample evidence that active labor market programs (ALMPs) played an important role in achieving this goal for more general population groups (Card et al., 2017). However, it is unclear how informative this literature is for more recent refugees, as they tend to be less educated and more culturally distant from the native population compared to earlier refugee waves (Dustmann et al., 2017).

To this date, only a limited number of empirical studies in Europe analyzed the effectiveness of active labor market polices for the recent refugee wave, mainly from Scandinavia (Arendt and Bolvig, 2020; Arendt, 2022; Dahlberg et al., 2020), in addition to a study from Austria (Ortlieb et al., 2020). Overall, the results highlight that on-the-job training and further vocational training provide significant employment and income effects, while short-term programs focusing on skill assessment and career counseling also showed positive (albeit more modest) effects. The results also show that early language training increases employability of refugees. We add to this literature by examining the effectiveness of the four quantitatively most important German ALMPs for welfare recipients (Wapler et al., 2018) on the labor market success of refugees in Germany. Our analysis is based on rich and high-quality administrative data, allowing us to evaluate the effectiveness of the programs using inverse probability weighting based on the selection-on-observables assumption (Heckman and Robb, 1985).

Our sample is based on the full count of refugees with a residence permit who were unemployed and received welfare benefits on September 30th 2016. On the one hand, choosing this sampling date allows a sufficiently long observation period after the start of treatment such that causal treatment effects up to 33 months after treatment start can be estimated. On the other hand, this limits the number of available observations – as only about 600.000 individuals had a residence permit – which grants them access to welfare in Germany towards the end of 2016 (German Federal Office for Statistics, 2022).

Moreover, only about 124.000 refugees were officially registered as unemployed and received welfare in September 2016 (German Federal Employment Agency, 2021). To obtain a more homogenous sample, we restrict our attention to individuals who entered Germany between 2013 and 2016, further limiting

sample size. Moreover, most refugees were male such that observation numbers for women in ALMPs are too low to estimate program effects separately. Hence, we focus on men only in this paper. These and a few other necessary sample restrictions leave us with a sample of about 30,000 to 33,000 male individuals, depending on the ALMP considered. Our analysis shows that relatively short-term training – either by a service provider or in-firm – and longer term further vocational training increase employment and earnings, while a public employment program is ineffective. Additional analyses show that there is non-negligible effect heterogeneity in terms of individuals' duration of stay in Germany for participants of training schemes by providers. Moreover, sensitivity analyses show that our results are robust to unobserved confounding.

Next, Section 2 presents the German institutional setting and provides a short overview of the evaluated programs. Section 3 lays out our empirical strategy, provides details on the data and presents selected descriptive statistics. Section 4 presents the empirical analysis, while Section 5 concludes.

2. Institutional background
2.1. Institutional setting

In Germany, two types of unemployment benefits are available. Unemployment benefits I (UB I) cover unemployed workers who contributed to the unemployment insurance system for a minimum of one year within the last two years of employment. UB I-beneficiaries receive up to 67% (60%) of their last net wage as parents (childless persons), usually for a maximum duration of one year (Riphahn et al., 2013).

Unemployment benefits II (UB II) or simply *welfare benefit*, are tax-financed means-tested flat rate benefits covering persons capable of working[1] for a minimum of three hours per day, and whose household income falls short of the legal social minimum. So even persons who are working or receive UB I can be eligible to UB II if their household income is sufficiently low (Riphahn et al., 2013; Wapler et al., 2018). Most refugees who arrived between 2013 and 2016 and who have a legal residence permit in Germany are also eligible for welfare benefits as their household income is likely below the social minimum. Moreover, most of them have not worked long enough to be eligible for UB I benefits.

Welfare beneficiaries are obliged to make efforts to increase their employability to reduce their welfare dependency. Public employment offices, referred to as job centers, support welfare beneficiaries to this purpose through job search assistance and targeted placement in ALMPs (Wapler et al., 2018). To combat moral hazard, welfare beneficiaries refraining to participate in an ALMP may be subject to sanctions that can result in a reduction of welfare benefits for a period up to three months (van den Berg et al., 2014). The same sanctions apply for refugees assigned to participate in an ALMP.

2.2. Which programs do we evaluate?

The first two programs examined are part of the so-called Schemes for Activation and Integration (SAI). In Germany, SAIs are performed either by private training providers (Schemes by Providers or SP) or by employers in the form of In-Firm Training (IFT), also referred to as on-the-job training (Harrer et al., 2020).

SP are the most frequently used ALMPs for welfare recipients in Germany. They provide training whose goals include guiding into apprenticeships and work, identifying and reducing employment impediments, providing placement services into contributory employment, preparing for self-employment and stabilizing an employment take-up (Wapler et al., 2018). For example, *Guiding into Apprenticeships and into Work* is a type of SP that provides guidance on finding suitable job offers and writing resumes and application letters. Another type of SP focuses on detecting individual employment impediments, improving skills and providing information on suitable employment opportunities (Harrer et al., 2020). SP may also be combined with In-Firm Training for a period of up to 4 weeks. Moreover, SP focusing on teaching a professional occupation may not exceed a period of 8 weeks (Harrer et al., 2020). The median duration of SP in our sample is about 3 months.

While SP occur mainly in the form of classroom training, IFTs occur in establishments and are shorter in duration. At the median, IFTs last one month in our sample. IFT measures focus on gaining practical skills and applying them directly on the job. In this manner, they are comparable to unpaid internships (Harrer et al., 2020). A number of refugee-specific programs were additionally developed within the context of a SAI. However, the numbers of treated individuals in these programs are typically too low to be analyzed by themselves or cannot be properly identified in the data due to initial limitations after the introduction of these programs.

The third program considered is Further Vocational Training (FVT), whose objective is to enable employability through human capital accumulation. FVT include short-term programs providing professional and practical skills along with long-term programs (up to two years) providing a certified vocational training degree (Bernhard and Kruppe, 2012; Wapler et al., 2018). Refugees can also receive accreditation of their foreign vocational qualifications in the context of FVT (German Federal Employment Agency, 2013).

As FVT programs usually require longer and more intensive investments in human capital, participants have less time and resources to invest in job search compared to non-participants, creating a so-called "lock-in-effect". The lock-in-effect reflects a period of lower employment rates for participants compared to non-participants due to being "locked-in" in the training. This lock-in-effect is likely to occur only during program participation, which, on average, lasts about five months in our sample. Soon after training completion, participants'

employment should catch up and even exceed those of non-participants due to higher returns from human capital investments (Arendt, 2022; Card et al., 2017).

Lastly, we examine a public employment program referred to as One-Euro-Jobs (OEJs) targeting unemployed welfare recipients strongly detached from the labor market. OEJs are temporary (part-time) public sector jobs, typically lasting between three and twelve months, in which participants receive one to two Euros per hour that is not deducted from their welfare benefit (Harrer and Stockinger, 2021). The median duration of OEJs in our sample is three months. OEJs provide basic work activities for participants to increase long-term employability. Jobseekers should only be placed in a One-Euro-Job if none of the above-mentioned ALMPs is (yet) suitable (Hohmeyer and Wolff, 2018). All the above-mentioned programs, apart from the refugee-specific programs, were therefore pre-existing and extended to refugees. Access considerations according to the Federal Employment Agency are hence identical for refugees as non-refugees and include the necessary language skills.

Comparing treatment probabilities across programs, Tübbicke and Kasrin (2021) show that unemployed refugees have a higher probability of participating in SP and IFT but a lower participation probability in FVT and OEJ compared to non-refugees. These differences may be due to different employment impediments among refugees on the one hand and caseworker preferences or perceptions about program effectiveness on the other. Case workers assign participants to programs after weighing potential program benefits against costs.

Higher participation in SP and IFT relative to FVT could be due to their shorter duration and due to lower human capital required for successful participation. FVT programs are also more expensive than SP and IFT. On the other hand, OEJs are programs of 'last resort' aimed at the long term unemployed who are extremely distant to the labor market. Previous results for more general populations also show that participants tend to differ across these programs. For example, Harrer et al. (2020) show that IFT participants have more favorable background characteristics, leading to better labor market prospects than participants of SP in the absence of treatment. Hohmeyer and Wolff (2012) show that OEJ participants are expected to perform relatively poorly in the labor market without an intervention. These differences must be addressed in order to obtain unbiased estimates of treatment effects (see Section 3.1).

3. Empirical approach and descriptive statistics
3.1. Identification and estimation strategy

We aim to estimate the causal effects of a series of active labor market programs on participating refugees in Germany. In the notation of the potential

outcomes model by Roy (1951) and Rubin (1974), we want to estimate the average treatment effect on the treated (ATT), defined as

$$\tau_{ATT}^g = E\big[Y_g^1|D_g = 1\big] - E[Y_g^0|D_g = 1], \tag{1}$$

where D_g is a treatment indicator, taking on the value of one if the person received treatment $g \in \{SP, IFT, FVT \text{ and } OEJs\}$, and zero if not. While Y_g^1 refers to the outcome that is realized when an individual receives treatment g, Y_g^0 denotes the outcome that occurs when the same individual is not treated with g (but potentially other programs). Hence, the ATT defined by (1) measures the additional effect of program g relative to all other programs available. Due to the fundamental evaluation problem, Y_g^0 is unobservable for participants and hence, the second expectation in the equation above has to be imputed from data on non-participants.

To do so, we use inverse probability weighting (IPW) based on a selection-on-observables approach (Heckman and Robb, 1985). IPW re-weights non-participants based on their propensity score $P_g(X) = \Pr(D_g = 1 \,|X)$, i.e. the conditional treatment probability, in order to undo the selection into treatment. Under the selection-on-observables assumption, i.e. the assumption that the set of observed characteristics X includes all relevant factors that simultaneously govern the selection process and the outcome of interest, and the so-called common support assumption $P_g(X) < 1 \ \forall \ X$, the missing counterfactual can be recovered from data on non-participants by reweighting their outcomes by the odds of treatment $P_g(X)/\big(1 - P_g(X)\big)$.

While the selection-on-observables assumption is strong, we are equipped with administrative data that allows us to control for a large number of observed characteristics measured prior to the start of treatment. To specify a model for $P_g(X)$, we largely follow other studies evaluating the programs of interest for more general populations and pick covariates that have been shown to be predictive of at least one of the treatments under consideration (Bernhard and Kruppe, 2012; Harrer et al., 2020; Harrer and Stockinger, 2021).

We adjust for socio-demographics (age, time in Germany, nationality, schooling and occupational qualification[2], presence of children by age category), household characteristics (household type, total income, income without welfare benefits), the type of last job (regular or minor employment) along with minor employment at sampling date, individuals' labor market history[3] (cumulated durations in regular employment, vocational training, ALMP participation, unemployment benefit I and welfare receipt, job search, information on regular employment, and real labor earnings at specific dates one year prior to the sampling date) and regional labor market information (unemployment rate, long-term unemployment rate, vacancy-to-unemployment ratio, inflow rates into sanctions and ALMPs).

Studies by Brücker et al. (2020) and Kosyakova et al. (2021) show that age, time in Germany, schooling, the presence of children in the household and regional labor market conditions are important predictors of refugees' labor market success. To minimize the impact of any parametric assumptions on our analysis, we discretized all covariates into sensible categories.[4]

If the set of covariates X does not include all factors that influence the treatment decision and the outcome, then selection-on-observables assumption will fail and estimates will be biased. For example, this could be the case if physical or mental health – characteristics we do not observe – are decisive for selection into treatment and the outcome-generating process. Similarly, bias may arise if unobserved language skills are important predictors of treatment and outcomes. Fortunately, (Caliendo et al., 2017) have shown that the inclusion of typically unobserved covariates tends not to significantly alter effect estimates of ALMPs upon conditioning on a large set of observed covariates. This is likely because covariates and especially pre-treatment outcomes have already been affected by those unobserved factors before the start of treatment.

Moreover, our list of control variables includes several strong predictors that can proxy language skills, including age, time since entry, education and country of origin (Fennelly and Palasz, 2003; van Tubergen, 2010). Nonetheless, we inspect how sensitive our results are to unobserved confounding in Section 4.4. Results shows that our estimates are highly robust. Hence, we believe that the selection-on-observables assumption is a reasonable approximation for our application. Because we have fairly low treatment to control ratios and differences in covariate distribution are modest, overlap between groups is strong and we do not need to trim the sample.[5]

In order to estimate the propensity score, we use the so-called covariate balancing propensity scores (CBPS, Imai/Ratkovic, 2014). The CBPS does not maximize the maximum-likelihood, instead it estimates logit coefficients by algorithmically optimizing the resulting covariate balance, leading to (near) perfect finite-sample balance.[6] In addition to obviating the need for iterative balance checking and re-estimation of the propensity score, it was shown to perform well in a recent empirical Monte-Carlo simulation by Frölich et al. (2017) based on data similar to ours. Denoting person i's estimated propensity score by $\hat{P}_{gi}(X)$ and its observed labor market outcome by Y_i, estimates of the ATT are obtained using the normalized IPW estimator as

$$\hat{\tau}^g_{ATT} = \frac{1}{\sum_{i=1}^N D_{gi}} \sum_{i=1}^N D_{gi} Y_i - \frac{1}{\sum_{i=1}^N \frac{(1-D_{gi})\hat{P}_{gi}}{1-\hat{P}_{gi}}} \sum_{i=1}^N \frac{(1-D_{gi})\hat{P}_{gi} Y_i}{1-\hat{P}_{gi}} \ . \qquad (2)$$

Intuitively, IPW re-weights non-participants by increasing (decreasing) the weight of individuals with a large (small) propensity score, thereby eliminating bias due to differences in covariate distributions between treated and untreated

units. For statistical inference, we rely on the weight-based asymptotic approximation of standard errors by Lechner (2001) as it was shown to be somewhat conservative and to perform rather well in large samples (see, Bodory et al., 2020, for details).[7]

3.2. Data and descriptive statistics

As a data source, we mainly use the Integrated Employment Biographies (IEB), an administrative dataset from the statistics department of the Federal Employment Agency in Germany. The data includes the universe of employees liable to social security, the registered unemployed, registered jobseekers, benefit recipients, and ALMP program participants. Using the IEB has a number of advantages. First, due to its extensive nature, the IEB allow to condition on a large number of observed characteristics. Second, as the IEB includes the universe of person groups, it allows observing the full count of individuals, eliminating the need to rely on a random sample as is the case for most survey data. Lastly, the IEB does not suffer from panel attrition, which is likely to be more problematic among refugees than among the general population. All of these factors should greatly improve inference in our application.

We use the full count of refugees with a residence permit according to §§ 22–26 Residence Law ("Aufenthaltsgesetz") who were unemployed and on welfare as of September 30, 2016. As the great majority of refugees arriving to Germany between 2013 and 2016 was male and their treatment probabilities are higher than those of female refugees, we restrict our analysis to men.[8] To get a more homogenous sample, we restrict our attention to individuals who first entered Germany in 2013 or later.[9] Moreover, we exclude individuals who were younger than 17 or older than 62 at the sampling date to avoid including people still in high school or who have reached the legal retirement age before the end of our analysis time frame. This is necessary as neither being in school nor in retirement can be observed in the data.

Based on this sample, we classify individuals as treated if they entered SP, IFT, FVT or OEJs during the entry window between October 2016 and March 2017.[10] For each treatment group we define a separate control group (or simply non-participants hereafter), i.e. individuals who did not participate in the respective program during the entry window. However, non-participants may participate in other programs during the entry window or participate later in the same program. Using a six months entry window allows sufficient observation numbers among the treated, while avoiding selectivity issues as a longer entry window may lead to a negative selection of non-participants (Sianesi, 2004). For each treatment, non-participants are randomly assigned a hypothetical entry month from the distribution of entry months among the respective treatment group. Individuals who are either gainfully employed or participating in one of the other programs at the hypothetical entry date are dropped from the

respective sample.[11] This leads to slightly different groups of non-participants across treatments.

Table 1 shows selected descriptive statistics for the four estimation samples. SP is clearly the most common ALMP assigned to refugees with 6,118 participants. Moreover, 747 individuals participated in IFT, followed by 465 OEJ and 424 FVT participants. The groups of non-participants range in size from around 27,000 to 30,000 individuals.

Panel A shows that the majority of both participants and non-participants are Syrian. Refugees from Iraq, Eritrea or Afghanistan make up between one and six percent, depending on the sample.[12] The mean age ranges from just over 30 years among participants of SP to around 33 years among individuals in FVT, while non-participants are about 32 years old. ALMP participants also differ by their time of arrival to Germany: on the one hand, OEJ and SP participants have been in Germany for 1.33 and 1.38 years, respectively. Both are significantly less than the 1.41 years that their non-participants spent in Germany. On the other hand, with an average duration of 1.67 and 1.9 years for IFT and FVT participants, respectively, both groups have spent more time in Germany than their non-participants.

Moreover, FVT and IFT participants have the highest percentage of persons with at least some vocational training or university education, with 33 percent for FVT and 16 percent for IFT participants. SP and OEJ participants have a lower share of persons with some professional qualification, at 10 percent for SP participants and 7 percent for OEJ participants. Among non-participants, 11 percent have at least some vocational training or university education. Similarly, IFT and FVT participants have the highest high school completion rates with 62 and 74 percent across all groups. Among OEJ participants, 53 percent completed high school, which is lower than the completion rates of non-participants with 59 percent. This indicates that IFT and FVT participants might have more favorable labor market prospects than SP participants, while OEJ participants seem to have the lowest employment prospects.

Our outcomes of interest are monthly regular employment, i.e. an indicator for whether an individual is in unsubsidized dependent employment subject to social security contributions, and real monthly labor earnings. For our causal analysis, we make use of outcomes from 12 months prior to the start of (hypothetical) treatment up to 33 months after. After 33 months, IFT and FVT participants perform best in comparison to the non-participants in both dimensions, as shown in panel B of Table 1. However, due to the non-random selection into treatment, such a simple mean comparison is plagued by selection bias.

4. Empirical analysis

In this section, we present estimates of causal effects for the ALMPs considered and inspect heterogeneity along several standard background characteristics.

TABLE 1. Selected descriptive statistics

	Schemes by providers		In-firm training		Further voc. Training		One-Euro-Jobs	
	Part.	Non-Part.	Part.	Non-Part.	Part.	Non-Part.	Part.	Non-Part.
A. Covariates								
Mean in years								
Age	30.5	32.0***	30.2	31.9***	32.8	31.8**	32.4	31.8
Time in Germany	1.38	1.41***	1.67	1.40***	1.90	1.40***	1.33	1.41**
High school diploma								
No/no information	0.39	0.41***	0.39	0.41	0.26	0.41***	0.47	0.41***
Yes	0.61	0.59***	0.62	0.59	0.74	0.59***	0.53	0.59***
Professional qualification								
None/no information	0.90	0.89**	0.84	0.89***	0.67	0.89***	0.93	0.89***
At least some vocational or university education	0.10	0.11**	0.16	0.11***	0.33	0.11***	0.07	0.11***
Share from ...								
Syria	0.83	0.83	0.82	0.83	0.83	0.83	0.87	0.83**
Iraq	0.05	0.06*	0.04	0.06***	0.02	0.06***	0.04	0.06**
Eritrea	0.05	0.04***	0.06	0.04***	0.06	0.04***	0.06	0.04**
Afghanistan	0.02	0.02	0.02	0.02	0.03	0.02**	0.01	0.02
Other	0.05	0.05	0.07	0.05	0.06	0.05	0.02	0.06***
B. Outcomes after 33 months								
Share in regular employment	0.33	0.29***	0.55	0.29***	0.50	0.29***	0.28	0.30
Mean monthly labor earnings	561	485***	947	442***	1413	474***	453	486
Number of obs.	6,118	27,135	747	28,942	424	29,903	465	29,805

Note: Statistically significant differences at the 10/5/1% level between participants and non-participants are indicated by */**/*** on the respective value among the non-participants.

TABLE 2. Balancing quality before and after weighting

	Schemes by providers		In-firm training		Further voc. Training		One-Euro-Jobs	
	Before	After	Before	After	Before	After	Before	After
Absolute standardized bias								
Mean in %	4.26	0.00	10.14	0.02	15.35	0.08	9.40	0.05
(Re-)estimation of the Propensity score								
Pseudo-R^2	0.05	0.00	0.09	0.00	0.20	0.00	0.17	0.00
p-value	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
Rubin's B	56.9	1.08	95.8	3.7	161.5	8.1	132.5	6.1
Number of obs.								
Participants	6,118		747		424		465	
Non-participants	27,135		28,942		29,903		29,805	

Note: The absolute standardized bias is defined as the difference in means between treatment and control divided by the square root of the mean of variances prior to weighting in both groups. Rubin's B is the standardized mean difference of the linear index of the propensity score of treatment and control group.

Lastly, a sensitivity analysis examines the robustness of estimates to unobserved confounding.

4.1. Checking balance

Before turning to the causal analysis, we present some evidence on the balancing quality before and after re-weighting the respective group of non-participants using our IPW approach. Table 2 displays several common balancing measures given by the mean absolute standardized bias (MSB, Rosenbaum and Rubin, 1983), the pseudo-R^2 from a probit regression, along with its corresponding p-value of overall significance (Sianesi, 2004), and Rubin's B (Rubin, 2001). If the CBPS methodology works, MSB, pseudo-R^2 and Rubin's B should be very close to zero after re-weighting. Indeed, all three measures only show negligible deviations from zero after weighting, indicating (near) perfect covariate balance. Similarly, the p-value rises from essentially zero to one in all samples after weighting, thus it is impossible to reject the null hypothesis that covariates are not related to treatment. Hence, our balancing approach delivers excellent balance, successfully purging the outcome gap between participants and non-participants from its association with background characteristics.[13]

4.2. Main results

Figure 1 presents effects on regular employment, with effects on real monthly labor earnings displayed in Figure 2. Estimated effects range from twelve months prior to the start of treatment until 33 months after. Estimates prior to treatment
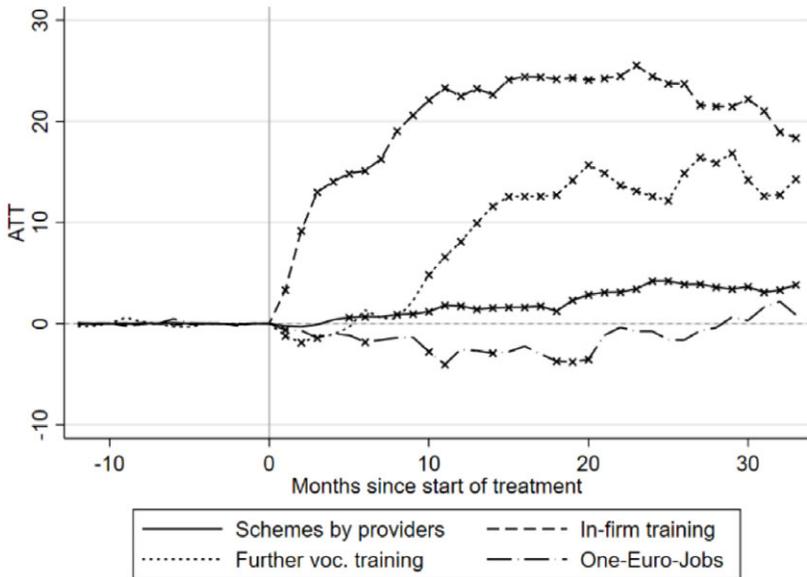
FIGURE 1. Effects on regular employment in percentage points.
Note: Statistically significant estimates at the 5% level are marked with a cross.
N(Participants): 6,118 for Schemes by Providers; 747 for In-Firm Training; 424 for Further Vocational Training; 465 for One-Euro-Jobs.
N(non-Participants): range from 27,135 for Schemes by Providers to 29,903 for Further Vocational Training.

are close to zero and statistically insignificant, showing that participants and non-participants follow the same path in terms of employment and earnings, providing additional evidence that our weighting strategy was successful.

For OEJs, there is a rather long-term and statistically significant negative effect on labor earnings and employment. After about 30 months, point estimates become positive but insignificant. Actually, this result is not surprising: OEJs are a last resort measure for individuals who are particularly distant from the labor market and so are likely to only increase employment prospects in the long run, possibly also in combination with other ALMP measures (Kiesel and Wolff, 2018). Indeed, the majority of OEJ participants in our sample received some additional treatment after the OEJ program. This partly explains the long-lasting negative effects of OEJ participation despite its median duration of around three months. Similar results are found by Harrer and Stockinger (2021) who analyze the effectiveness of OEJs for a more general population of unemployed welfare recipients.

SP show a lock-in effect for the first three months after the start of the program, which coincides with their median duration in our sample. After five months, employment effects become statistically significant and after seven months, earnings are positively affected by the treatment. After 33 months,
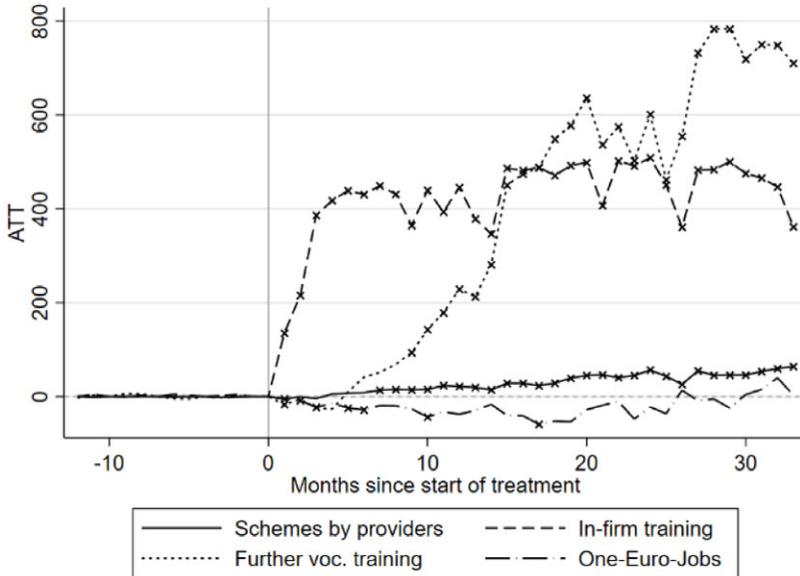
FIGURE 2. Effects on real monthly labor earnings in Euro.
Note: Statistically significant estimates at the 5% level are marked with a cross.
N(Participants): 6,118 for Schemes by Providers; 747 for In-Firm Training; 424 for Further Vocational Training; 465 for One-Euro-Jobs.
N(non-Participants): range from 27,135 for Schemes by Providers to 29,903 for Further Vocational Training.

employment effects are roughly 3.8 percentage points and effects on earnings are about 65 Euro. While in absolute value, these effects may seem rather small, they are considerable when compared to mean outcomes of participants (see Table 1). Relative to the mean, these effects are equivalent to an 11.5 percent increase in employment and a 13.4 percent increase in monthly labor earnings.

For IFT, we find positive effects on participants' employment probabilities almost immediately. This highlights that participating in an IFT, which has a medium duration of one month, speeds up labor market integration considerably. Over time, employment effects increase, peaking at around 25 percentage points after 23 months. Afterwards, estimates of employment effects subside somewhat, reaching 18 percentage points at the end of the observation period.[14] Effects on earnings follow a similar path, with an impact on monthly labor earnings of 360 Euro after 33 months. The effect sizes for IFT and SP participants are comparable to those found for a more general population of unemployed welfare recipients by Harrer et al. (2020), who examined treatment effects of program entry into SP and IFT in 2010.

The results for FVT show a significant lock-in effect for three months, which is not surprising given its relatively long median duration of 5 months.

After around 10 months, positive impacts on employment rates and average earnings materialize. While employment effects quickly stabilize at around 14 percentage points, effects on earnings rise to around 710 Euro after 33 months. Again, the effect sizes are similar to the ones found by Bernhard and Kruppe (2012) for German and non-German welfare participants who entered FVT in 2005.

Thus, our findings indicate that all programs except OEJ are effective in improving refugees' labor market success. As groups of program participants differ in their composition, it should be noted that effect sizes across programs cannot be directly compared.

4.3. Effect heterogeneity

For our heterogeneity analysis, we split our samples according to median values of certain background characteristics, given by age, time in Germany and regional unemployment. The analysis then repeats the estimation steps for all sub-samples. As before, the CBPS methodology yields excellent balancing quality after weighting (see Table A.1 in the appendix for details).

First, we split the sample according to individuals' age, i.e. under 30 or at least 30 years old. This distinction is important because younger refugees might learn German and adapt to their new environment quicker than older refugees, which may influence how much they benefit from training programs. Second, we split the sample by duration of stay in Germany since time of arrival; here we analyze effects for individuals with a short and a long duration, i.e. below and above the median of roughly 1.4 years. Along this dimension, those with a longer duration are expected to be better integrated socially and culturally – and thus may benefit more from programs targeting short-term labor market integration such as SP and IFT.

We also analyze effects for individuals in different local labor market environments, splitting the sample by the local unemployment rate, with 9 percent serving as the cut-off. Refugees living in regions with relatively high unemployment may profit the most from programs focusing on human capital acquisition such as FVT, due to higher skill mismatch in regions with higher unemployment (Wapler et al., 2018). Table 3 shows estimates of employment effects for all programs and sub-groups to keep the presentation concise.[15]

While all subgroup effect estimates are statistically significant except for OEJs, only SP shows statistically significant effect heterogeneity for duration of stay in Germany (indicated by the italics in Table 3). Specifically, the employment probability of persons with a longer duration increases by 5.5 percentage points through participation, more than twice its impact on persons with a shorter duration of stay. Although heterogeneity effects for all other estimates are statistically insignificant, gaps are sizable among IFT, FVT and OEJ participants along several dimensions. As these are relatively small treatment groups, a

TABLE 3. Heterogeneous effects regarding regular employment in percentage points

	Schemes by providers (SP)	In-firm training (IFT)	Further voc. Training (FVT)	One-Euro-Jobs (OEJ)
Age				
Under 30	3.6***	16.8***	14.2***	0.7
	(3,363; 13,335)	(410; 14,291)	(173; 14,945)	(220; 14,854)
At least 30	3.9***	20.0***	15.0***	1.0
	(2,755; 13,800)	(337; 14,651)	(251; 14,958)	(245; 14,951)
Time in Germany				
Short	2.3**	14.2***	20.0***	2.9
	(3,194; 13,385)	(213; 14,512)	(77; 14,909)	(240; 14,821)
Long	5.5****	20.1***	12.7***	−1.7
	(2,924; 13,749)	(534; 14,430)	(347; 14,994)	(225; 14,984)
Unemployment rate				
Low	4.2***	17.5***	13.5***	3.5
	(3,269; 15,731)	(498; 16,631)	(222; 17,250)	(200; 17,198)
High	3.3***	21.2***	13.9***	−2.5
	(2,849; 11,404)	(249; 12,311)	(202; 12,653)	(265; 12,607)

Note: Statistically significant estimates at the 10/5/1% level are marked with */**/***. Italics indicate statistically significant differences at the 5% level between group-specific ATTs. Numbers of observations (treated; comparison) are in brackets below the estimates.

lack of statistical power makes the detection of meaningful effect heterogeneity difficult. Particularly interesting are the results found for IFT and OEJs. Similar to SP, IFT participants with longer durations of stay appear to benefit more compared to participants with shorter durations. Regarding OEJs, point estimates suggest that this type of program may be effective for individuals with a short duration of stay in Germany and for individuals in regions with low unemployment rates. The latter may be due to an easier transition from a OEJ to a regular job in low unemployment environments.

4.4. Sensitivity analysis

The results presented so far are based on the selection-on-observables assumption. Hence, it is crucial to assess the risk of those estimates being plagued by "hidden bias" through unobserved confounding (Rosenbaum, 2002). As we cannot rule out this possibility, we follow much of the literature on sensitivity analyses and explore how point estimates and possibly inference may change under certain assumptions regarding a relevant but omitted binary confounder.

Vanderweele and Arah (2011) show that – under some simplifying assumptions – the bias B_g due to an unmeasured binary confounder U can be written as:

TABLE 4. Sensitivity analysis

	Schemes by providers (SP)	In-firm training (IFT)	Further voc. Training (FVT)	One-Euro-Jobs (OEJ)
Baseline estimates	3.8***	18.3***	14.3***	0.9
Distribution of estimates taking potential unobserved confounders into account				
Minimum	3.1***	16.7***	11.0***	−0.1
5th percentile	3.6***	17.9***	13.6***	0.7
Median	3.9***	18.3***	14.3***	0.9
95th percentile	4.0***	18.5***	14.9***	1.2
Maximum	4.6***	18.9***	15.8***	1.7
Number of obs.				
Participants	6,118	747	424	465
Non-participants	27,135	28,942	29,903	29,805

Note: Statistically significant estimates at the 10/5/1% level are marked with */**/***. This table shows estimation results using the sensitivity analysis suggested by Vanderweele and Arah (2011). The analysis assumes that there is some unobserved covariate that is distributed the same as an observed covariate and re-estimates bias-adjusted causal effects.

$$B_g = \delta_y \{\Pr(U = 1|D_g = 1) - \Pr(U = 1|D_g = 0)\}, \tag{3}$$

where δ_y is the population average effect of switching the unobserved confounder U from zero to one. Similar to Ichino et al. (2008), we use the distribution of observed covariates to inform our analysis about plausible potential unobserved confounder distributions. To provide the most credible results possible, the entire list of available covariates is used. The bias components in Formula (3) are estimated as follows: estimates of δ_y are obtained through linear regression of the employment indicator after 33 months on all available covariates among non-participants. The sample shares of covariates among participants and non-participants serve as estimates of their probabilities of possessing the respective characteristic. Based on this information, the true ATT is re-estimated under the assumption that there exists an unmeasured confounder, which follows the same distribution as the observed covariate. Table 4 shows our baseline estimates as presented in Section 4.2 along with information on the distribution of results for the sensitivity analyses.

As expected, there is some variation in effect estimates, albeit rather limited considering the strong predictiveness of covariates such as age, unemployment and job search history along with regional unemployment on individuals' employment probabilities.[16] Moreover, statistical inference remains unchanged, independent of the type of unobserved confounder assumed. This lack of variation in the bias-adjusted estimates supports our claim that the selection-on-observables

assumption seems reasonable. If there is some bias remaining after conditioning on observed characteristics, it appears unlikely to overturn our inference.

5. Conclusion

This paper provides first evidence on the effectiveness of four prominent ALMPs on the labor market success of recent refugees in Germany based on a rich administrative dataset. Using the universe of male refugees on welfare that were unemployed on September 30, 2016 and entered treatment between October 2016 and March 2017, we examine treatment effects for a period of around three years after program entry. Our empirical strategy relies on the selection-on-observables assumption, which is justifiable in light of the high-quality administrative data and supported by our sensitivity analysis.

We find that all programs, except for the public employment scheme, have a positive impact on participants' regular employment and monthly earnings in the medium run. Thus, our analysis confirms that ALMPs are a valuable tool for successfully integrating male refugees into the German labor market. Our results on the positive effects of In-Firm Training and Further Vocational Training on regular employment are in-line with the results found for Sweden by Dahlberg et al. (2020). Moreover, our findings on the positive employment effects of In-Firm Training are consistent with the results found for Denmark by Arendt (2022) and Arendt and Bolvig (2020).

In contrast to Arendt and Bolvig (2020), who find that In-Firm Training only increases employment in the short-term, our results show steady employment effects after around 3 years of program entry. Additionally, we find a significantly positive effect of In-Firm Training on wages in both the short and the medium term while the results by Arendt (2022) show no significant short-term effect. This implies that ALMPs are an important addition to other existing European integration programs such as language courses that were also found to be effective.

Another interesting result is that the absolute effect sizes found are similar to those shown for other more general populations of unemployed welfare recipients in Germany (see Harrer et al., 2020; Bernhard and Kruppe, 2012; Harrer and Stockinger, 2021). Yet, refugees typically face stronger employment impediments, meaning that their effects are even more economically significant. This is because migrants in general, and refugees in specific, tend to display lower labor market success than natives, even after controlling for important factors including age and education (Algan et al., 2010; Bedaso, 2021). A possible reason for the larger relative effect for refugees is that successful ALMP participation acts as a stronger positive signal to potential employers due to their larger distance from the labor market compared to other groups (Liechti et al., 2017).

Regarding policy recommendations, the results clearly imply that public employment schemes should be used less intensively for refugees. A viable alternative may be to intensify the use of other ALMP programs for refugees, such as start-up and wage subsidies, which were shown to be effective for the general population as well as migrants (Butschek and Walter, 2014; Wolff et al., 2016). This may be a more effective option for individuals more distant from the labor market than public employment schemes, as start-up and wage subsidies typically lead to more regular types of employment, generating more valuable labor market experience. Lastly, one may think of combining several types of ALMPs more frequently.

However, both issues, i.e. start-up and wage subsidies as well as the combination of different ALMPs needs to be examined in future research for refugees to make definite recommendations. Analyses should also be performed for women when the number of observations allows it, to provide a comprehensive picture on the effectiveness of ALMPs on the successful integration of refugees in the German labor market.

Competing interests

The author(s) declare none.

Notes

1 Persons unable to work due to sickness, disability or old age are entitled to social assistance.
2   Schooling and occupational qualification are measured using the most recent information available. For most refugees, this coincides with their schooling and qualification prior to migration as they typically have not had the chance to upgrade their human capital since arrival. Occupational qualifications are measured by having at least some vocational training or university experience as most refugees do not have a formal degree.
3 The IEB only contains the labor market history of refugees during their time in Germany. One may worry that this could cause estimates to be less reliable compared to studies focusing on the more general population where the labor market history of people is highly predictive of their future labor market success. However, Kosyakova et al. (2021) show that labor market experience before the transition to Germany is only weakly statistically predictive for refugees' employment. Hence, this concern does not seem critical.
4 For example, including age in linear and squared form in the estimation would implicitly assume that balancing the first and second moment is sufficient to remove bias. To avoid this, we inspected covariate distributions of continuous variables and formed reasonably sized categories of at least 30 observations each.
5 This was shown by inspecting the propensity score distributions, both visually and using summary statistics. To be concise, we refrain from reporting these results. We have also

    implemented more strict common support measures and results do not change in any meaningful way compared to our baseline results.

6 Balancing is only perfect up to the tolerance given by the optimization procedure, meaning some negligible imbalance remains even after re-weighting.

7 While results by Bodory et al. (2020) show that bootstrapping techniques may outperform the Lechner (2001) approximation, the large number of replications necessary, our relatively large dataset and the computational intensity of the CBPS method makes the bootstrap infeasible in our application.

8 On September 30th 2016, about 124.000 people registered refugees were unemployed and received welfare, about 90,000 of which were men. Dropping individuals who entered Germany before 2013 and those whose date of entry is unknown reduces the sample to about 37.400 male individuals. As the date of entry into Germany is unknown in job centers run solely by municipalities, only individuals who are administered by job centers jointly administered by the municipality as well as the Federal Employment Agency are kept in the sample. In addition to the age restriction, individuals who left welfare, took up employment or joined another ALMP between sampling and (hypothetical) program start are dropped to implement the estimation procedure. Final numbers of observations are shown in Table 1. Due to lower observation numbers and lower treatment probabilities for women, an analysis including women cannot compare effects of different ALMPs. However, estimated effects of SP participation for women are presented in Kasrin et al. (2021).

9 Another reason is that this makes our sample as comparable as possible to the sample from the IAB-BAMF-SOEP survey (which also includes information on refugees entering between 2013 and 2016), allowing us to use studies based on the survey to understand which individual and household characteristics influence labor market success and employment quality for a population similar to our population of interest. Restricting our analysis to sample members who arrived in Germany in 2015 or later (80% of the original sample) barely changes point estimates but inflates standard errors noticeably.

10 Treated individuals may receive another treatment after the ones we analyze. Indeed, this happens quite frequently: between 31 percent (SP) and 59 percent (OEJ) receive another treatment after the start of the treatments analyzed. However, due to relatively small sample sizes, it is impossible to analyze such interaction effects. Nonetheless, our sensitivity analyses show that multiple treatments do not drive our main results.

11 Since treatments are evaluated independently of one another, an individual in the group of non-participants for one treatment may be part of the treatment group for another treatment. Moreover, individuals in the group of non-participants may join a different or the same program after the entry window. Overall, between 54 percent (IFT) and 57 percent (SP) of non-participants ever receive some treatment in the period of observation. However, our sensitivity analyses show that such a contamination of the control groups does not drive our results.

12 As 80% of the sample are Syrian and due to the limited number of participants in most programs, we were unable to perform subgroup analyses by nationality. We did however estimate effects for Syrians only and found similar results to those found for the total sample.

13 An analysis of the resulting weight distribution implied by the CBPS method shows that no trimming is necessary in order to avoid overly large weights as proposed by Imbens (2004).

14 This drop is hardly statistically significant at the 5% level. Additional inspection shows that the drop is driven by participants' employment rates stagnating after 23 months while individuals from the control group catch up to some degree.

15 Patterns of effect heterogeneity for real monthly labor earnings are similar and found in Table A.2 in the Appendix.

16 Certain categories of these covariates have regression coefficients of 20 percentage points or above.

References

Algan, Y., Dustmann, C., Glitz, A. and Manning, A. (2010), The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom. *The Economic Journal*, 120, 4–30.

Arendt, J. N. (2022), Labor market effects of a work-first policy for refugees. *Journal of Population Economics*, 35(1), 169–196.

Arendt, J. N. and Bolvig, I. (2020), *Early labor market entry, language acquisition and labor market success of refugees. Vive Working Paper*, 11274.

Bedaso, F. (2021), *The Labor Market Integration of Refugees and other Migrants in Germany, GLO Discussion Paper*, 884, Global Labor Organization (GLO), Essen.

Bernhard, S. and Kruppe, T. (2012), Effectiveness of Further Vocational Training in Germany–Empirical Findings for Persons Receiving Means-tested Unemployment Benefits. *Journal of Applied Social Science Studies*, 132(4), 501–526.

Bodory, H., Camponovo, L., Huber, M. and Lechner, M. (2020), The finite sample performance of inference methods for propensity score matching and weighting estimators. *Journal of Business & Economic Statistics*, 38(1), 183–200.

Brücker, H., Kosyakova, Y. and Schuß, E. (2020), Fünf Jahre seit der Fluchtmigration 2015: Integration in Arbeitsmarkt und Bildungssystem macht weitere Fortschritte. IAB-Kurzbericht, 04/2020.

Butschek, S., Walter, T. (2014), What active labour market programmes work for immigrants in Europe? A meta-analysis of the evaluation literature. *IZA Journal of Migration* 3, 48.

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

Card, D., Kluve, J. and Weber, A. (2017), What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3), 894–931.

Clark, A. E. (2003), Unemployment as a social norm: Psychological evidence from panel data. *Journal of labor economics*, 21(2), 323–351.

Dahlberg, M., Egebark, J., Vikman, U. and Özcan, G. (2020), Labor Market Integration of Low-Educated Refugees: RCT Evidence from an Ambitious Integration Program in Sweden. IFN Working Paper, 1372.

Dustmann, C., Fasani, F., Frattini, T., Minale, L. and Schönberg, U. (2017), On the economics and politics of refugee migration. *Economic Policy*, 32(91), 497–550.

Fennelly, K. and Palasz, N. (2003), English language proficiency of immigrants and refugees in the Twin Cities metropolitan area. *International Migration*, 41(5), 93–125.

Frölich, M., Huber, M. and Wiesenfarth, M. (2017), The finite sample performance of semi- and non-parametric estimators for treatment effects and policy evaluation. *Computational Statistics & Data Analysis*, 115, 91–102.

German Federal Employment Agency (2013), Fachliche Hinweise „Förderung der beruflichen Weiterbildung nach § 16 Abs. 1 SGB II i.V.m. §§ 81 ff. SGB III" (March 2013).

German Federal Employment Agency (2021), *Personen im Kontext von Fluchtmigration – Deutschland, Länder, Kreise, Agenturen für Arbeit und Jobcenter (Monatszahlen, September)*.

German Federal Office for Migration and Refugees (2021), *Das Bundesamt in Zahlen 2020: Asyl, Migration und Integration*.

German Federal Office for Statistics (2022), *Ausländer: Deutschland, Stichtag, Geschlecht, Aufenthaltstitel, Staatsangehörigkeit. Table 12521-0008*.

Harrer, T., Moczall, A. and Wolff, J. (2020), Free, free, set them free? Are programmes effective that allow job centres considerable freedom to choose the exact design?. *International Journal of Social Welfare*, 29(2), 154–167.

Harrer, T. and Stockinger, B. (2021), First step and last resort: One-Euro-Jobs after the reform. *Journal of Social Policy*, 51(2), 412–434.

Heckman, J. J. and Robb Jr, R. (1985), Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1-2), 239–267.

Hohmeyer, K. and Wolff, J. (2012), *Arbeitsgelegenheiten im SGB II: Zwei Varianten mit unterschiedlicher Wirkung*. IAB-Kurzbericht, 9/2012.

Hohmeyer, K. and Wolff, J. (2018), Of carrots and sticks: the effect of workfare announcements on the job search behaviour and reservation wage of welfare recipients. *Journal for Labour Market Research*, 52(1), 1–23.

Ichino, A., Mealli, F. and Nannicini, T. (2008), From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity?. *Journal of applied econometrics*, 23(3), 305–327.

Imai, K. and Ratkovic, M. (2014), Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 243–263.

Kasrin, Z., Stockinger, B. and Tübbicke, S. (2021), *Aktive Arbeitsmarktpolitik für arbeitslose Geflüchtete im SGB II: Der Großteil der Maßnahmen erhöht den Arbeitsmarkterfolg*. IAB-Kurzbericht, 7/2021.

Kiesel, M. and Wolff, J. (2018), *Langfristige Teilnahmewirkungen von Ein-Euro-Jobs: Das Einsatzfeld hat Einfluss auf die Integrationschancen*. IAB-Kurzbericht, 8/2018.

Klinger, S. and Fuchs, J. (2020), Effects of population changes on the labour market in Germany. *IAB-Forum*, 01/04/2020.

Kosyakova, Y., Gundacker, L., Salikutluk, Z. and Trübswetter, P. (2021), *Arbeitsmarktintegration in Deutschland: Geflüchtete Frauen müssen viele Hindernisse überwinden*. IAB-Kurzbericht, 08/2021.

Lechner, M. (2001), Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric evaluation of labour market policies*. Heidelberg: Physica, 43-58.

Liechti, F., Fossati, F., Bonoli, G. and Auer, D. (2017), The signalling value of labour market programmes. *European sociological review*, 33(2), 257–274.

Ortlieb, R., Eggenhofer-Rehart, P., Leitner, S., Hosner, R. and Landesmann, M. (2020), Do Austrian Programmes Facilitate Labour Market Integration of Refugees?. *International Migration*, 1-21.

Paul, K. I. and Moser, K. (2009), Unemployment impairs mental health: Meta-analyses. *Journal of Vocational behavior*, 74(3), 264–282.

Pohlan, L. (2019), Unemployment and social exclusion. *Journal of Economic Behavior & Organization*, (164/August 2019), 273-299.

Riphahn, R. T., Sander, M. and Wunder, C. (2013), The welfare use of immigrants and natives in Germany: the case of Turkish immigrants. *International Journal of Manpower*, 34(1), 70–82.

Rosenbaum, P. R. (2002), Overt bias in observational studies. In *Observational studies*. Springer, New York, NY, 71–104.

Rosenbaum, P. R. and Rubin, D. B. (1983), The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.

Roy, A. D. (1951), Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2), 135–146.

Rubin, D. B. (1974), Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701.

Rubin, D. B. (2001), Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3), 169–188.

Sianesi, B. (2004), An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics*, 86(1), 133–155.

Tübbicke, S. and Kasrin, Z. (2021), Aktive Arbeitsmarktpolitik erhöht den Arbeitsmarkterfolg von Geflüchteten teilweise deutlich (Serie "Befunde aus der IAB-Grundsicherungsforschung 2017 bis 2020"). IAB-Forum, 15.07.2021.

van den Berg, G. J., Uhlendorff, A. and Wolff, J. (2014), Sanctions for young welfare recipients. *Nordic Economic Policy Review*, (1/2014), 177-208.

van Tubergen, F. (2010), Determinants of second language proficiency among refugees in the Netherlands. *Social Forces*, 89(2), 515–534.

Vanderweele, T. J. and Arah, O. A. (2011), Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 42-52.

Wapler, R., Wolf, K. and Wolff, J. (2018), *Do active labour market policies for welfare recipients in Germany raise their regional outflow into work? A matching function approach*. IAB-Discussion Paper, 8/2018.

Wolff, J., Nivorozhkin, A. and Bernhard, S. (2016), You can go your own way! The long-term effectiveness of a self-employment programme for welfare recipients in Germany. *International Journal of Social Welfare*, 25(2), 136–148.

Appendix

TABLE A1. Covariate Balancing before and after weighting – heterogeneity analysis

	Schemes by providers		In-firm training		Further voc. training		One-Euro-Jobs	
	Before	After	Before	After	Before	After	Before	After
Age								
Under 30	4.41	0.01	10.28	0.03	15.50	0.47	10.47	0.21
	(3,363; 13,335)		(410; 14,291)		(173; 14,945)		(220; 14,854)	
At least 30	3.76	0.02	10.81	0.07	15.9	0.07	9.86	0.08
	(2,755; 13,800)		(337; 14,651)		(251; 14,958)		(245; 14,951)	
Time in Germany								
Short	4.38	0.13	8.76	0.23	13.9	0.15	9.12	0.05
	(3,194; 13,386)		(213; 14,512)		(77; 14,909)		(240; 14,821)	
Long	4.38	0.00	7.91	0.15	12.35	0.04	11.01	0.08
	(2,924; 13,749)		(534; 14,430)		(347; 14,994)		(225; 14,984)	
Unemployment rate								
Low	3.76	0.00	10.43	0.14	16.69	0.27	12.32	0.25
	(3,269; 15,731)		(498; 16,631)		(222; 17,250)		(200; 17,198)	
High	6.65	0.03	10.80	0.05	15.66	0.44	9.82	0.09
	(2,849; 11,404)		(249; 12,311)		(202; 12,653)		(265; 12,607)	

Note: This table shows the mean absolute mean bias (MSB) in percent for before and after weighting for all sub-samples of the heterogeneity analysis. The absolute standardized bias is defined as the difference in means between treatment and control divided by the square root of the mean of variances prior to weighting in both groups. Numbers of observations (treated; comparison) are in brackets below the MSB.

TABLE A2. Heterogeneous effects regarding monthly labor earnings in Euro

	Schemes by providers (SP)	In-firm training (IFT)	Further voc. Training (FVT)	One-Euro-Jobs (OEJ)
Age				
Under 30	57.5***	304.7***	851.0***	16.7
	(3,363; 13,335)	(410; 14,291)	(173; 14,945)	(220; 14,854)
At least 30	66.1***	414.3***	617.4***	16.3
	(2,755; 13,800)	(337; 14,651)	(251; 14,958)	(245; 14,951)
Time in Germany				
Short	38.2***	227.1***	1095.1***	13.3
	(3,194; 13,385)	(213; 14,512)	(77; 14,909)	(240; 14,821)
Long	96.0***	431.2***	618.3***	7.7
	(2,924; 13,749)	(534; 14,430)	(347; 14,994)	(225; 14,984)
Unemployment rate				
Low	85.4***	302.4***	910.0***	43.8
	(3,269; 15,731)	(498; 16,631)	(222; 17,250)	(200; 17,198)
High	40.0**	495.4***	483.0***	42.5
	(2,849; 11,404)	(249; 12,311)	(202; 12,653)	(265; 12,607)

Note: Statistically significant estimates at the 10/5/1% level are marked with */**/***. Italics indicate statistically significant differences at the 5% level between group-specific ATTs. Numbers of observations (treated; comparison) are in brackets below the estimates.