

An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan

*Stefano Caria, Grant Gordon, Maximilian Kasy, Simon Quinn, Soha Shami,
Alexander Teytelboym*

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan

Abstract

We introduce a novel methodology for adaptive targeted experiments. Our Tempered Thompson Algorithm balances the goals of maximizing the precision of treatment effect estimates and maximizing the welfare of experimental participants. A hierarchical Bayesian model allows us to adaptively target treatments at different groups. We implement our methodology in a field experiment. We examine the impact of three interventions designed to improve formal employment outcomes of Syrian refugees and local jobseekers in Jordan: one treatment to address liquidity constraints, one to address information frictions, and one to address challenges of self-control. Six weeks after being offered treatment, none of the interventions has a significant or meaningful impact on the probability that individuals are in wage employment; we estimate that our targeting algorithm had a positive but small effect on aggregate employment (approximately 1 percentage point). However, we find large employment effects of all treatments for refugees at the two-month follow-up, and suggestive evidence of four-month impacts for the cash grant; liquidity appears to be a key barrier to employment for refugees.

JEL-Codes: C93, J6, O15.

Keywords: adaptive experiments, refugees, job search.

Stefano Caria
University of Warwick
stefano.caria@warwick.ac.uk

Grant Gordon
International Rescue Committee
grant.gordon@rescue.org

Maximilian Kasy
University of Oxford
maximilian.kasy@economics.ox.ac.uk

Simon Quinn
University of Oxford
simon.quinn@economics.ox.ac.uk

Soha Shami
Danish Refugee Council
soha.s.shami@gmail.com

Alexander Teytelboym
University of Oxford
alexander.teytelboym@economics.ox.ac.uk

Version of August, 2020.

Acknowledgments: We are grateful to David McKenzie, Rachael Meager, Magne Mogstad, Aleksey Tetenov and Eva Vivalt for helpful and stimulating comments.

1 Introduction

Randomized controlled trials (RCTs) have become a widely used method for policy evaluation (Duflo and Banerjee, 2017). In a conventional RCT, the designer randomly assigns treatments to experimental subjects in order to precisely estimate the effects of all treatments. In many contexts, however, the experimenter is not merely interested in learning whether policies work. Instead, the experimenter wants to maximize the welfare of program participants. To do so, the experimenter only needs to learn which treatment works best. If the experimenter observes treatment outcomes over time, she can use this information in order to adaptively optimize treatment assignment for future experimental participants.

Our first contribution is to introduce a methodology for adaptive targeted experimentation that balances competing goals of precise treatment effect estimation and maximizing the benefits to experimental participants. Our Bayesian algorithm has two key features. First, it is adaptive, i.e., it changes treatment assignment probabilities over time by incorporating information about the successes of treatments of existing experimental participants. Second, it is targeted, i.e., it uses information about the success rates of treatments in every group in order to target treatments for each individual group.

Our second contribution is to implement our methodology in a field experiment. As far as we know, ours is the first implementation of adaptive targeting in a field experiment in development economics. Our field experiment tested three active labour market policies for Syrian refugees and local workers in Jordan. We targeted treatments at 16 different strata of refugees and local workers. We find that our treatments have had minimal impact on six-week employment outcomes of jobseekers. We also find that there have been modest gains from targeting.

Tempered Thompson Algorithm within a hierarchical Bayesian model The first key feature of our methodology is that our treatment assignment is adaptive. The problem of adaptively assigning treatments in order to maximize outcomes during the experiment is known as a *multi-armed bandit* (MAB) problem (Scott, 2010). MAB problems are often

computationally intractable and a large literature in statistics has been devoted to finding tractable and effective heuristics to solve them. But MAB heuristics pose a problem for an experimenter interested in estimating the effects of all treatments: if the experimenter is quickly convinced that a particular treatment is *suboptimal*, she should stop assigning it in the future. As a result, the experimenter might miss out on learning about the effectiveness of good, though suboptimal, policies.

Our Tempered Thompson Algorithm combines the estimation objective of conventional RCTs with the welfare-maximizing objective of bandit algorithms. The designer starts with a prior over the effectiveness of k different treatments; we recommend a diffuse and symmetric default prior. Every period, the designer observes the outcomes of some of the current participants in the experiment. As a result, the designer can estimate the posterior probability \hat{p}_t^{dx} that treatment d is optimal for individuals from stratum x at time t . Then, at time t , the Tempered Thompson Algorithm assigns treatments in the following way, for individuals from stratum x :

With probability γ : assign treatment d to individual i with probability $\frac{1}{k}$.

With probability $(1 - \gamma)$: assign treatment d to individual i with probability \hat{p}_t^{dx} .

The Tempered Thompson Algorithm generalizes two classical treatment assignment protocols. When $\gamma = 1$, our algorithm boils down to a conventional randomized controlled trial. When $\gamma = 0$, our algorithm is the [Thompson \(1933\)](#) algorithm used in many online contexts, including platform revenue management, movie recommendations, and ad placement ([Russo et al., 2018](#)). However, when $0 < \gamma < 1$, the Tempered Thompson Algorithm (asymptotically) maximizes welfare of the participants subject to the constraint that every treatment has a probability of assignment at least $\frac{\gamma}{k}$. This allows the designer to target participant welfare while ensuring that they can learn something about the effectiveness of suboptimal treatments. Our main theoretical result ([Theorem 1](#)) formally establishes a tradeoff between the welfare of participants and the precision of the estimates: as γ increases, the expected variance of treatment effect estimators falls, but the expected outcomes of participants also decrease.

The second key feature of our methodology is that our adaptive assignment algorithm

is targeted. We use a hierarchical Bayesian approach in order to estimate heterogeneous treatment effects. Our model allows us to learn the extent of effect heterogeneity across different strata. At each time period t , the treatment effect of each treatment d in each stratum x is estimated as a weighted average of the observed success rate for d in x and the observed success rates for d across all other strata. The weights are determined optimally by the observed amount of heterogeneity across all strata as well as the available sample size in a given stratum. The posterior probability \hat{p}_t^{dx} that treatment d is optimal for stratum x is then calculated from this posterior distribution of treatment effects.

Implementation and Results We implement our methodology in a field experiment designed to help Syrian refugees and local jobseekers in Jordan find formal wage work. The experimental design and empirical analysis were specified before the start of the experiment in a pre-analysis plan submitted to the AEA registry.¹ The field experiment tests three types of support: a small, unconditional cash transfer (worth about 20 percent of average monthly expenditure); information provision to increase the ability to signal skills to employers; and psychological support to strengthen job search motivation. These types of support correspond to three barriers — material, informational, and behavioral — that refugees and locals might face in finding and retaining jobs. The program was implemented in Jordan by the International Rescue Committee at the height of the Syrian refugee crisis. Jordan is a relevant context in which to study employment policies for refugees, for at least two reasons. First, employment generation for refugees is a pressing policy concern in Jordan. In Jordan, an estimated 63% of refugees are unemployed and over 90% of Syrian refugees live below the national poverty line (Verme et al., 2015). The massive influx of unemployed, impoverished refugees into Jordan mirrors the type of displacement shock countries often experience. Second, and in response to the displacement crisis, the international community and Government of Jordan launched the Jordan Compact, the legal framework for refugees to access those jobs. In exchange for preferential access to the European market and access to conditional financing, the Government of Jordan agreed to provide 200,000 work permits for refugees. The Jordan Compact has influenced refugee policy around the world and similar compacts are being launched in other countries, for example Ethiopia. Jordan thus provided an opportune context to understand how to connect refugees to the new employment opportunities that are opening for them. Ours is the

¹ Available at <https://doi.org/10.1257/rct.3870-2.2>.

first field experiment to study the employment of refugees in a development context.

In the experiment, we set $\gamma = 0.2$ in the Tempered Thompson Algorithm to ensure that in every period every one of three treatments and the control has at least 0.05 probability of being assigned. We define 16 strata: $\{\text{Syrian, Jordanian}\} \times \{\text{Female, Male}\} \times \{\text{High school, No high school}\} \times \{\text{Never employed, Ever employed}\}$. Program intake started in mid February 2019 and ended in December 2019. Overall, we sampled 3,770 individuals, approximately evenly split between Syrians and Jordanians. We track participants' employment outcomes with a short phone interview six weeks after treatment, which we use to implement our Tempered Thompson Algorithm. Further, we carry out full phone surveys two and four months after treatment. These surveys enable us to measure a broader set of impacts and to study effects over a longer time period.²

Our first finding is that, six weeks after being offered treatment, none of the interventions has significant or meaningful impact on the probability that individuals are in wage employment (the primary outcome that we specified in our pre-analysis plan). However, while the control-treatment difference in outcomes is close to zero, we estimate that the average impact of the optimised treatment (i.e. of offering the best possible intervention to each stratum) is about a 1 percentage point increase in employment, suggesting some moderate short-term gains from targeting.

Second, we find that the cash intervention has large and significant impacts on refugee employment and earnings, two and four months after treatment. While employment rates remain stubbornly low in the control group, the cash grant raises job search rates and enables refugees to place more job applications. As a result, four months after treatment, the grant boosts employment by 3.8 percentage points (70 percent) and earnings by 65 percent. These are sizable impacts compared to those documented in the recent literature on active labor market policies (McKenzie, 2017). We also document substantial increases in hourly wages and in the probability of retaining a job between the two and four month interviews, indicating that match quality has also increased. Finally, consistent with the existence of binding liquidity constraints, we find that these impacts are driven by individuals with

² We were unable to complete a six month follow-up interview due to the national lockdown in Jordan during the Covid pandemic.

below-median expenditure a baseline, and that baseline expenditure is significantly associated with job search intensity in the control group.

Third, the information and psychological interventions also boost job search among refugees and have significant impacts on employment and earnings after two months. However, these impacts are smaller than those of the cash grant and are ultimately short lived. Four months after treatment, we find weaker and insignificant impacts of these interventions.

Fourth, we find essentially no positive effects of treatment on the Jordanian sample.³ While Jordanians and Syrians were sampled in a similar way and have identical baseline employment rates, Jordanians tend to be more educated and to have higher baseline expenditure. Further, control Jordanians search at much higher intensity than control Syrians and have better employment outcomes two months after baseline. This group may thus face weaker or different job search frictions, which are not addressed by our interventions.

These results shed light, for the first time, on the barriers to employment opportunities faced by refugees in a developing-country context. This evidence is particularly relevant for policy, as governments around the world consider expanding legal access to labor markets for refugees. In particular, our results point to the key role played by liquidity constraints, as in classical models of poverty traps ([Banerjee and Newman, 1993](#); [Balboni et al., 2020](#)). Our comprehensive findings on these constraints — including the large employment impacts of a small unconditional cash grants, a strong control association between liquidity and job search intensity, and the large heterogeneity of treatment effects with respect to liquidity — represent some of the cleanest evidence in the recent experimental literature on

³ We find a large and significant impact of job search from the psychological intervention, but no impacts on employment or earnings. Further, the cash and information interventions do not have significant impacts on any of our pre-specified outcomes.

the impacts of limited liquidity in urban labor markets.⁴ At the same time, the job-search impacts of the other two interventions, which do not provide additional liquidity, show that cash is not a binding constraint for all refugees in our sample. Both information and motivation seem to further limit participation in labor markets.

Related literature Our paper spans two distinct literatures. Methodologically, our work is related to experimentation, MAB problems, and targeted treatment assignment. While there is a large theoretical literature on optimal experimentation in MAB problems (e.g., [Gittins \(1979\)](#)), the bedrock of our analysis is “probability matching” algorithm due to ([Thompson, 1933](#)). Recently, a number of papers have shown that the Thompson algorithm asymptotically matches the welfare under the optimal dynamic treatment assignment policy ([Agrawal and Goyal, 2012](#); [Kaufmann et al., 2012](#); [Agrawal and Goyal, 2013](#)). We contribute to a growing number of papers in economics using adaptive experimental methods ([Kasy and Sautmann, 2019](#); [Kasy and Teytelboym, 2020a,b](#)). There is also a recent literature within economics on targeted treatment assignment both from a non-Bayesian (e.g., [Kitagawa and Tetenov \(2018\)](#); [Wager and Athey \(2018\)](#)) and Bayesian perspectives (e.g., [Dehejia \(2005\)](#); [Chamberlain \(2011\)](#); [Kasy \(2018\)](#)).

We also contribute to the literature on active labour market policies in developing and emerging economies. Specifically, ours is the first field experiment on employment of refugees in a development context.⁵ The literature on active labour market policies has generally found that such policies have limited effectiveness ([McKenzie, 2017](#)). This includes three novel experiments among educated youth in Jordan: one involving wage subsidy vouchers ([Groh et al., 2016a](#)), one involving training in soft skills ([Groh et al., 2016b, 2015](#)), and one involving direct matching of job-seekers to firms ([Groh et al., 2015](#)). However,

⁴ There is consistent evidence that interventions that provide much larger cash grants, worth up to one year of income, do not discourage work ([Banerjee et al., 2017](#)). These interventions typically aim at fostering entrepreneurship, are often evaluated in rural contexts, and do not measure impact on job search. Our findings are unique in that they study the effect of a much smaller grant and identify impacts on job search in an urban labor market. The only comparable study is [Banerjee and Sequeira \(2020\)](#), who find a small unconditional cash grant boosts job search, but not employment, among young South Africans. Further, other studies such as [Abebe et al. \(2020\)](#) analyse the impacts of conditional transfers that simultaneously relax individuals’ budget constraint and decrease the ‘price’ of job search relative to other types of consumption. Thus, they do not offer direct evidence on binding liquidity constraints.

⁵ [Battisti et al. \(2019\)](#) evaluate a job-matching intervention for recently-arrived refugees in Germany.

in other contexts, recent experiments have identified several effective policy interventions: conditional cash transfers have been found to increase short-term employment through increasing job search ([Franklin, 2018](#); [Abebe et al., 2020](#); [Banerjee and Sequeira, 2020](#)), skill-signalling workshops can increase wages through improved assortative matching ([Alfonsi et al., 2020](#); [Bassi and Nansamba, 2020](#); [Abebe et al., 2020](#)), and detailed job-search plans have increased employment through more effective job search ([Abel et al., 2019](#)). We draw on each of these three recent areas of innovation to design our three treatments. The previous literature tends to focus on young nationals with poor attachment to the labour market (see, for example, [Kluve et al. \(2019\)](#)). Our work is novel in taking insights from those earlier experiments to a population of refugees, for whom constraints may be quite different. In this way, our paper also relates to recent attempts to generalize experimental results across different contexts (see, for example, [Meager \(2019\)](#)).

Roadmap The paper is organized as follows. Section 2 describes the humanitarian and the labour market context in Jordan, our sampling procedure, and the three treatments. Section 3 explains our adaptive treatment assignment algorithm and derives its theoretical properties. Section 4 presents our empirical findings, including qualitative evidence from focus group interviews. Section 5 is a conclusion. Appendix A.1 gives the proof of the main theorem. Appendix A.2 provides details on the Markov Chain Monte Carlo algorithm for the hierarchical Bayesian model. The Online Appendix contains treatment materials used in the field as well as additional tables and figures.

2 Context, sampling and treatments

The world is facing the largest refugee crisis since World War II, with over 70 million individuals displaced, about 25 million of whom are refugees ([UNHCR, 2019a](#)). Amidst this crisis, the duration of displacement has increased with refugees now displaced for 10 years on average ([Devictor and Do, 2017](#)). The unprecedented magnitude and changing nature of displacement has catalyzed a radical shift in thinking about how assistance is provided for refugees and internally displaced people.

Over the past decade, the international community has moved away from a model in which refugees are housed in camps – receiving aid in perpetuity – to a model focused on identifying sustainable solutions that integrate refugees and IDPs into local communities and labor markets. In many contexts, this has fueled a change from delivering basic commodities and food items to supporting individuals to gain access to employment. This change in approach is not isolated to any specific location, but is increasingly becoming the dominant model for delivering humanitarian assistance.

A crucial part of integrating displaced individuals into labor markets is providing the support necessary to generate employment opportunities at scale for communities affected by crises. However, there is a dearth of evidence on what works for these groups and in these contexts. In part, this is due to the challenging nature of experimenting in crisis-affected contexts – where security issues and the need to deliver timely services make experimentation difficult. More generally, refugees and internally displaced individuals face a unique set of constraints in accessing employment opportunities. They often lack the information, language skills and social networks needed to effectively navigate labor markets. Many have lost assets and have limited savings; this can constrain individuals from accessing the type of childcare, transit or basic needs required to get a job. Trauma, uncertainty and social exclusion may also reduce refugees’ intrinsic motivation to search for an employment opportunity. These micro-level barriers may be compounded at the national level by governments who impose legal restrictions on whether or what types of jobs can be accessed.

2.1 The Syrian refugee crisis

Since 2012, the Syrian crisis has displaced more than 13.1 million people, making it the largest refugee crisis of our time ([UNHCR, 2020](#)). Approximately seven million are displaced internally within Syria; about another six million fled to neighbouring countries. The Government of Jordan estimates that, since the beginning of the Syrian crisis, nearly 1.3 million refugees have arrived in the country; of these, about 660,000 have registered with UNHCR ([UNHCR, 2020](#)). Eight years into the conflict, Syrian refugees in Jordan face important needs for humanitarian assistance, for basic services, and for economic stability. Today, it is estimated that 93% of Syrian refugees in the country live below the US\$5 per

day poverty line. At the same time, low-skilled Jordanians continue to suffer from pre-existing labor market challenges, including high-unemployment, which leaves them also economically vulnerable ([IRC, 2017](#); [Government of Jordan, 2019](#); [UNHCR, 2020](#)).

In an attempt to address some of the issues associated with the protracted displacement, the Government of Jordan and the international community met at the London Conference in 2016 and explored new ways to support countries most affected by the Syrian crisis. For Jordan, a key outcome of the event was the signing of the Jordan Compact — hailed at the time as an innovative approach for host countries and the international community to respond to protracted displacement. Under the Compact, European and international donors pledged a total of US\$2.1 billion in direct grants and US\$1.9 billion in concessional loans to the Government of Jordan ([Barbelet et al., 2018](#)). The Compact also granted Jordan trade concessions that relaxed ‘rules of origin’ criteria and opened export markets in Europe. In exchange, the Government of Jordan committed to important policy changes aimed at drawing Syrian refugees into the labor market. Among these changes are ([IRC, 2017](#)):

1. Easing administrative procedures to allow Syrian refugees to apply for work permits in the sectors open to employing them, namely manufacturing, agriculture, and construction – with a goal of providing work permits for up to 200,000 Syrian refugees;
2. Designating and developing five industrial zones, later called the Special Economic Zones (SEZs), that would be provided with maximum investment and trade incentives under the new investment law;
3. Allowing Syrian refugees to formalize existing businesses and to set up new businesses; and
4. Providing a small percentage of contractual Syrian employment opportunities in municipal works.

The impetus for this breakthrough agreement was that policies that eased access to European markets were expected to lead to higher demand for Jordanian exports, which in turn would create new jobs and boost formal employment for both refugees and Jordanians, mainly in the manufacturing sector and within the SEZs. In short, the Compact aimed to turn ‘the Syrian refugee crisis into a development opportunity’ ([Government of Jordan, 2016](#)).

2.2 The Jordanian labor market

The labour market in Jordan is characterised by very low employment rates, by international standards. For example, the Employment and Unemployment Survey (EUS) reports, for the last quarter of 2016, an employment rate of 30 percent and overall labor force participation rate of 36 percent.⁶ This very low average masks significant heterogeneity by gender. Among males, labor force participation is close to 59 percent, while among females it drops to 13.5 percent. [Fallah et al. \(2019\)](#) compile EUS figures for a longer period of time, showing that some of these are persistent features of the Jordanian labor market.

Employment rates among refugees are much lower than among Jordanians. In early 2017, the Jordan Labor Market Panel Survey (JLMPS) was adapted to include an almost-representative sample of Syrian refugees in Jordan. According to the JLMPS figures, the employment rate among Syrian refugees stood at 14 percent. Among women refugees, the employment rate dropped to 2 percent. This employment was often informal and median monthly salaries were below the national minimum wage.⁷ These figures are broadly consistent with the number of work permits issued under the Jordan compact. Of the targeted 200,000 work permits to be issued to Syrian refugees by 2020, 159,000 had been issued as of the end of 2019 ([UNHCR, 2019b](#)). However, this figure includes permits for jobs that have been terminated; it is likely that active permits are a much lower number. For example, according to some estimates, about 40,000 permits were active in May 2017 (out of a refugee population of more than 600,000) ([DSP and Columbia, 2020](#)).

Employment among Syrian refugees is likely to be constrained by both demand and supply side factors. On the labour demand side, firms often report difficulties in processing work permits for Syrians but also fear the consequences of sanctions applied to informal

⁶ The labor force participation rates gives the ratio of economically active individuals (employed or looking for work) over total working-age individuals in the country.

⁷ 75 percent of refugees reported that they did not have a formal work contract. This is most likely an underestimate of the rate of informality, as many refugees may be reluctant to report informal work. In the same questionnaire, 99 percent of refugees reported that their employer was not making social security contributions – a key indicator of formality. In terms of salaries, the median monthly salary was 187 JOD, while the formal minimum wage was 200 JOD.

work.⁸ Further, refugees face strong competition from both Jordanian nationals and other migrants. This is partly because firms are required to meet a quota of employing at least 15% Jordanians. Moreover, migrant workers (mostly from South Asia) were established and employed in large numbers in many of the low-paying jobs that were opened to Syrians as part of the Compact ([Amjad et al., 2017](#)).

On the labor supply side, several search frictions are likely to be present. First, refugees are often credit-constrained due to lost assets, networks, and sources of income ([Government of Jordan, 2019](#)). Second, they have little experience in and information on the formal labor market in the host economy, which could drive decisions to work informally or not work at all. Third, they may experience substantial self-control problems when it comes to searching for work, possibly resulting from the psychological pressures of displacement and/or a number of restrictive labor market policies ([Shami, 2019](#)). Lastly, job quality in the formal sector is often a barrier to labour supply. Recent evidence shows that both Syrians and Jordanians perceive that formal work, particularly in the manufacturing sector, is often exhausting, exploitative, and potentially exposing to risk ([Amjad et al., 2017](#); [Razzaz, 2017](#)).

2.3 Sampling Syrian and Jordanian job-seekers

Our study sample enrolled in the IRC’s Project Match on a rolling basis over a six-month period between February 10, 2019 and November 30, 2019. The program was active in three cities: the capital Amman, and the northern cities of Irbid, and Mafraq. To be eligible for this study, participants had to be: (i) Syrian refugees or Jordanian nationals with valid government identification, (ii) between 18 and 45 years old (inclusive), and (iii) willing to take up low-skilled formal wage work that pays approximately minimum wage (220 JODs per month) in the immediate future. We verified that the participants met these requirements and further collected information on participants for the research during the intake registration interviews. At the end of the interview, participants were then randomized into a treatment group based on the algorithm described in section 3.

⁸ In particular, Article 12 of the Jordanian Labor Law identifies three violations to employing Syrian refugees: “(i) employing a non-Jordanian without a work permit; (ii) a non-Jordanian working for an employer other than one approved by the Ministry of Labour; and (iii) a non-Jordanian working in a profession other than the one approved by the Ministry of Labour” ([Amjad et al., 2017](#)).

Participants were selected using a variety of passive and active recruitment methods. The passive methods entailed IRC employment service officers (ESOs) contacting potential program participants. We refer to this as ‘passive’ selection as it was initiated by the ESO and not by the program participant. In the majority of cases, employment officers learned about potential program participants from referrals given by community leaders, other programs or partner organizations, and other study participants. Additionally, the ESOs conducted door-to-door home visits to neighborhoods that were known to host a high number of refugees. These neighborhoods were identified using UNHCR maps and the experience of ESOs hired to work with Project Match. Further, individuals who had not been contacted by an ESO were also eligible to apply for the program. We refer to this as ‘active’ selection as it was initiated by the program participant. Individuals could enrol by visiting specific community-based organizations (CBOs), visiting to IRC offices, responding to ads posted on social media, or by attending an information session on Project Match at the UNHCR offices.

There were no major differences in the way Syrians and Jordanians were sampled. For both Syrians and Jordanians, the largest share of enrolments came from referrals, a passive sampling method. The second largest source of participants for both nationalities was enrollment by the job-seeker at a CBO (an active sampling method). Slightly more Syrians than Jordanians were sampled through home visits conducted by the ESOs. However, overall, low-skilled and more economically vulnerable Jordanian often resided in areas similar to those of refugees and also engaged actively with CBOs to access various forms of welfare. We summarise the frequency of these different sampling methods by nationality in Table A.1 in the Online Appendix.

The proportion of participants enrolled through passive versus active methods changed over time, but not dramatically. In particular, in the months of May to July, 2019, more participants enrolled in Project Match through active methods. In subsequent months, this was largely reversed. We illustrate these patterns in Figure A.1 of the Online Appendix.

Table 1: **Descriptive statistics**

Sample	All	Syrian	Jordanian
	(1)	(2)	(3)
Female	0.60	0.60	0.60
Age	28.82	29.66	28.15
Household head	0.27	0.38	0.19
Household size	4.88	4.98	4.80
Educaton (years)	10.24	7.71	12.24
Spent at least 5 years in Jordan	-	0.95	-
Wage employed	0.02	0.02	0.02
Work experience (years)	4.48	4.99	4.10
Sample size	3770	1663	2107

2.4 Key features of the sample

In total, we sampled 1,663 Syrians and 2,107 Jordanians. We report a battery of descriptive statistics in Table 1. On several dimensions, the Syrian and Jordanian samples have similar characteristics. For both nationalities, 60 percent of the sample is composed by women, average age is about 29 years, and the average household is composed of about 5 individuals. Also, 2 percent of individuals of both nationalities are in wage employment and the average person has 5 years of work experience. Syrians however tend to be much less educated on average (7 years vs 12 years).

We divide this sample in sixteen strata based on four dummy variables: (i) nationality (a dummy for whether the respondent is Jordanian, defined. as having a Jordanian national ID); gender (a dummy for being female), (iii) education (a dummy for having completed high school or more), and (iv) work experience (a dummy for having experience in wage employment). These strata will form the basis of our targeting strategy, discussed in the next section. In Figure A.2 of the Online Appendix, we show the distribution of observations across strata. While for most cells we have good sample sizes, we tend to have a small proportion of people, especially Syrians, that have some education beyond high school.

An important point to stress is that many individuals in our sample, including the refugees, are actively looking for work; about 40 percent of refugees in the control group are doing so at the time of our one-month follow-up interview.

2.5 Treatments

On the basis of these key features, and working closely with local experts at the International Rescue Committee in Amman, we designed three separate job search interventions.⁹ Each intervention was designed to represent a distinct form of job search assistance, each having support in the recent empirical literature.¹⁰ Search interventions are aimed at facilitating the job search and thereby, increasing job search intensity to improve chances of participants finding work. These interventions will be denoted by $D \in \{0, 1, 2, 3\}$ where 0 refers to respondents assigned to a control group; the three search interventions respectively provide cash, information, and psychological support. In addition to these treatments, all respondents received 4 Jordanian dinars ('JOD': about US\$5.60 USD at the time of the intervention) to cover possible costs of transport to a job interview, and an informational flyer covering steps for interview preparation.¹¹

Control group: The control group received the 4 JODs and informational flyer that were offered to everyone upon registration with Project Match. Additionally, they received continuous case management conducted by trained employment service officers (ESOs) over the course of six months. During the follow-up calls, ESOs collected information for research purposes and they also responded to job-related concerns whenever possible.

Treatment 1: A labeled cash transfer. The cash support is a labeled cash transfer (LCT) of a value of 65 JOD (about US\$92 at the time of the intervention). This transfer was intended to support the recipient to pay for the cost of job search – including transport, grooming,

⁹ We prototyped and modified the interventions with about 130 respondents before commencing the randomized field experiment.

¹⁰ Some respondents were also assigned to one of two separate 'direct placement' arms; this is the focus of a separate paper.

¹¹ This was done to encourage participants to enrol in Project Match and to partially address potential ethical concerns of randomization by offering a placebo to the control group.

time costs and, for at least some study participants, childcare. It was designed based on evidence that small transfers cause large responses in job-search intensity ([Herkenhoff et al., 2016](#); [Franklin, 2018](#); [Abebe et al., 2020](#)). The transfer was ‘labeled’ in that, at the time of distribution, study participants were offered recommendations on how they should use this cash, i.e., to help with the job search in the above-mentioned ways); however, respondents were also informed that there would be no enforcement of whether the cash was actually used in this way ([Benhassine et al., 2015](#)). Upon delivery of the intervention, participants received an empty ATM card, which was charged (within an average of seven working days) with a one-time cash payment of 65 JOD. Upon charging of the ATM card, recipients receive an SMS notification. They also receive an ATM guide pamphlet with a direct hotline number for reporting issues with cash withdrawal from ATMs.

Treatment 2: Information. The second intervention provided informational support. Prior evidence suggested that both Syrian refugees and low-skilled Jordanians had little understanding of either the interview process or the legal obligations owed by employers to their workers ([Gordon, 2017](#)). (For example, a common myth among Syrian refugees in Jordan is that, by working in a formal job and holding a work permit, the Syrian would lose her or his UNHCR financial assistance package.¹²) Specifically, respondents in this treatment received information on (i) how to prepare for and interview for a formal job in Jordan (following, in particular, the recent results from [Abebe et al. \(2020\)](#)), and (ii) the legal rights of employees in formal jobs. Information was delivered through face-to-face interaction with a trained Project Match employment service officer (ESO), two videos describing the formal jobs and associated labor laws from the eyes of a job-seeker, and two take-home paper tools. The paper tools were designed for low-literacy participants and include cartoons for easy comprehension (see Online Appendix Figures [B.1](#) and [B.2](#)). One of the tools was designed as an interactive myth-busting activity whereby participants are exposed to common myths about formal jobs and worker rights, and then upon scratching the surface of the box below the myth, can see the reality.

Treatment 3: Psychological support. The third intervention is psychological support (which we refer to as the ‘nudge’ intervention). We provide a packaged intervention composed of

¹² The legal reality is that UNHCR financial assistance is not linked to having a work permit; instead, it depends upon a thorough financial needs assessment.

(i) a four-week job-search planning calendar as in (Abel et al., 2019) (see Online Appendix Figure B.3), (ii) an instructional video on how to use the calendar to plan for the job search, (iii) a face-to-face demonstration delivered by the ESOs, and finally (iv) reminder SMSs. The instructional video begins with a nudging statement of the potential impact of planning on employment from other contexts, ‘Did you know that job search planning can increase chances of finding work by up to 25%?’. Additionally, the reminder SMSs are given once at the beginning of the week and once at the end of the week to help respondents overcome self-control problems related to job search. Through the calendar and the SMSs, participants track the number of jobs and search hours they intend to apply for and spend respectively and then report back on the number of jobs and hours they actually apply for and spend that week. This intervention is motivated by recent evidence indicating substantial self-control problems and intention-behavior gaps in job search (DellaVigna and Paserman, 2005; Caliendo et al., 2015; Abel et al., 2019).

All interventions were delivered at the end of the intake interview or in the following seven days.

2.6 Follow-up surveys and attrition

We measure the impacts of these interventions through three follow-up surveys, all administered over the phone. We complete in-depth surveys two and four months after the baseline interview.¹³ We use these surveys to document the impacts of the program on a battery of outcomes specified in our pre-analysis plan.

We also complete a very short follow-up survey six weeks after baseline. This survey is focused exclusively on measuring whether the respondent is currently in wage employment. We use the data from this survey to implement the adaptive randomization design which we describe in the following section.

¹³ We also planned a six month follow-up survey, which we were unable to complete due to the COVID-19 crisis and the strict lockdown that was imposed in Jordan.

3 Treatment assignment and inference

In this section we describe our treatment assignment algorithm. Our algorithm is a modification of Thompson sampling (Thompson, 1933; Russo et al., 2018). This modification is motivated by the fact that our experiment has two objectives. Our primary objective is to get as many experimental participants into formal employment as possible. Our secondary objective is to test the effectiveness of alternative interventions.

We implement our algorithm within a hierarchical Bayesian model; cf. Gelman et al. (2014). The data-generating process for the binary potential outcomes corresponding to each treatment and stratum is governed by a parameter. For a given treatment, these parameters come from a common prior distribution for all strata. The hyper-parameters governing the common prior distribution are assumed to come from a diffuse hyper-prior distribution. The posterior distribution can be interpreted as follows. In every period, the experimenter observes treatment success rates for existing experimental participants across the strata, allowing her to learn the hyper-parameters. She can then combine the estimate of the hyper-parameters with the observed success rate in a given stratum in order to calculate the posterior distribution of the success parameter in that stratum. Finally, these posterior distributions can be used to calculate the probabilities \hat{p}_t^{dx} that a given treatment is optimal for a given stratum. These probabilities are then used in the Tempered Thompson Algorithm.

After describing this Bayesian setup, we review Thompson sampling. Thompson sampling is based on the posterior probability that each of the treatments is optimal, conditional on observed covariates. We then introduce our modification, the Tempered Thompson Algorithm, which provides a compromise between Thompson sampling and full (balanced) randomization. In Theorem 1 we characterize how the Tempered Thompson Algorithm trades off our two objectives, helping participants and obtaining precise estimates. The source code for our assignment algorithm is available in a public repository.¹⁴

This section concludes with a discussion of inference. Our primary method of inference

¹⁴ At <https://github.com/maxkasy/ThompsonHierarchicalApp>. A corresponding interactive app is available at <https://maxkasy.github.io/home/hierarchicalthompson/>.

is Bayesian. We also discuss p-values based on randomization inference, as a secondary method. The latter needs to take into account the adaptive and targeted form of treatment assignment in order to be valid.

We use the following notation. Let t denote the day of the intervention and let i index individuals within days. Note that we have repeated cross-sections, not a panel, so that individual i on day t is different from individual i on day t' when $t \neq t'$. Let x index strata and d index treatments. Finally, m_t^{dx} denotes the total number of times that treatment d was assigned to individuals in stratum x up to time t , and r_t^{dx} denotes the corresponding total number of successes, that is, individual for whom $Y_{it} = 1$.

3.1 Hierarchical Bayesian model

We consider a hierarchical Bayesian model with a data generating process, described by Eq. (1), and a prior, described by Eqs. (2) and (3) below. Let θ^{dx} be the average potential outcome for treatment d in stratum x . We assume that

$$Y_{it}^d | (X_{it} = x, \theta^{dx}, \alpha^d, \beta^d) \sim \text{Ber}(\theta^{dx}), \quad (1)$$

$$\theta^{dx} | (\alpha^d, \beta^d) \sim \text{Beta}(\alpha^d, \beta^d), \quad (2)$$

$$(\alpha^d, \beta^d) \sim \pi, \quad (3)$$

where (α^d, β^d) are the hyper-parameters and π is the hyper-prior (cf. [Gelman et al. \(2014, chapter 5\)](#)). Eq. (2) says that for a given treatment d , average potential outcomes θ^{dx} for all strata come from a Beta distribution governed by the hyper-parameters. Eq. (3) states that the hyper-parameters governing the distribution of average potential outcomes of each treatment across strata come from a common hyper-prior distributon π .

We assume that parameters $(\alpha^d, \beta^d, \theta^{d\cdot})$ are independent across the treatment arms d . We choose a hyper-prior for the hyper-parameters (α^d, β^d) with a common density equal to $(\alpha + \beta)^{-2.5}$, up to a multiplicative constant. In doing so, we follow the recommendation of [Gelman et al. \(2014, p.110\)](#) for picking a “non-informative” hyper-prior.

Intuitively, updating based on this prior works as follows. For each treatment d , we consider

the success rates $q_t^{dx} = r_t^{dx} / m_t^{dx}$ across the different strata x . Based on these success rates, we learn the mean and dispersion of θ^{dx} across strata, as reflected in hyper-parameters (α^d, β^d) . Then we use these as a prior, which together with the cumulative successes r_t^{dx} observed for a given stratum x allows us to form an updated belief about θ^{dx} for that stratum.

Denote by θ, m_t, r_t the vectors of parameters, cumulative trials, and cumulative successes, where each of these is indexed by both d and x , and denote by α, β the vectors of hyper-parameters indexed by d . We sample from the posterior distribution of (θ, α, β) given m_{t-1}, r_{t-1} using the Markov Chain Monte Carlo algorithm described in Algorithm 1 in Appendix A.2.

3.2 Treatment assignment algorithm

Let p_t^{dx} denote the posterior probability that a treatment d is optimal in stratum x , in the sense that it maximizes the probability of employment. That is, define

$$p_t^{dx} = P \left(d = \arg \max_{d'} \theta^{d'x} | m_t, r_t \right). \quad (4)$$

Equation (A.1) in the appendix shows how to estimate this probability by an average across Markov Chain Monte Carlo draws, which we denote \hat{p}^{dx} .

Two popular algorithms for assigning treatments in experiments are (i) fully random assignment, with equal probabilities across arms, and (ii) Thompson sampling. Our experiment is based on a combination of these two algorithms.

Fully randomized sampling assigns treatment d with probability $1/k$, where $k = 4$ is the number of different treatments, to units in every stratum. These assignment probabilities maximize power for tests of non-zero treatment effects. Thompson sampling, by contrast, assigns treatment d with probability \hat{p}_t^{dx} to units in stratum x in time period t . Thompson sampling minimizes expected regret (cf. [Agrawal and Goyal 2012](#); [Bubeck and Cesa-Bianchi 2012](#)), or equivalently maximizes average outcomes, in the large sample limit. As shown in these papers, it is in particular the case that expected regret only grows at a logarithmic rate with the number of experimental units. [Russo and Van Roy \(2016\)](#) prove worst-case bounds

on the performance of Thompson sampling, using information-theoretic arguments.

Our primary goal is to maximize the labor market outcomes of experimental participants, but we also consider the precision of treatment effect estimates to be a secondary objective. Motivated by this combination of objectives, we assign treatment d to units in stratum x with probability

$$(1 - \gamma) \cdot \hat{p}^{dx} + \gamma/k. \quad (5)$$

where γ is the share of observations that are randomized between treatment arms with equal probability. We will refer to this procedure as Tempered Thompson Algorithm sampling.

In our experiment, we measure employment outcomes Y_{it} only with a delay, six weeks after the intervention took place for each participant. As a consequence, treatment assignment is conditioned only on the outcomes of participants from six weeks before, or earlier. We assign participants in the first six weeks randomly to each treatment arm with probability 0.25.

3.3 Large sample properties

We now turn to a formal characterization of the large sample properties of our treatment assignment algorithm. We recapitulate and summarize our assumptions for this characterization in Assumption 1. In the following, we use θ_0 to denote the fixed, true vector of average potential outcomes from which the data are generated. By contrast, we use θ to denote the corresponding random vector which is drawn from the posterior distribution (belief) of the experimenter. The first step in Theorem 1 below, then, is based on the result that the posterior converges to the truth, that is, the distribution of θ concentrates around θ_0 .

Assumption 1 (Setup) *Consider a fixed (non-random) $\theta_0 = (\theta_0^{dx})$. Suppose that $d^{*x} = \arg \max_d \theta_0^{dx}$ is unique for all $x \in \{1, \dots, n_x\}$, and denote $\Delta^{dx} = \max_d \theta_0^{dx} - \theta_0^{dx}$. Assume that $(Y_{it}^1, \dots, Y_{it}^k, X_{it})$ is i.i.d. across both i and t , and that*

$$Y_{it}^d | (X_{it} = x, \theta_0) \sim \text{Ber}(\theta_0^{dx}).$$

Assume that $N_t \geq \underline{N}$ for all t and some constant \underline{N} , and that the prior distribution for θ has full support.

Assume that treatment d is assigned to units in stratum x in period t with probability

$$(1 - \gamma) \cdot p_t^{dx} + \gamma/k,$$

where p_t^{dx} equals the posterior probability that treatment d is optimal in stratum x , and $0 < \gamma \leq 1$. Denote q_t^{dx} the cumulative share of observations assigned to treatment d in stratum x across the time periods $1, \dots, t$, and p^x the probability that $X_{it} = x$.

Theorem 1 (Large sample properties of Tempered Thompson Algorithm) Under Assumption 1, the following holds true as t (and thus $M_t = \sum_{t' \leq t} N_{t'}$) goes to ∞ :

1. Consistency:

The posterior probability p_t^{dx} that treatment d is optimal in stratum x converges to 1 in probability (conditional on θ_0) for $d = d^{*x}$, and to 0 for all other d .¹⁵

2. Converging shares:

The cumulative share q_t^{dx} allocated to treatment d in stratum x converges in probability to $\bar{q}^{dx} = (1 - \gamma) + \gamma/k$ for $d = d^{*x}$, and to $\bar{q}^{dx} = \gamma/k$ for all other d .

3. Converging regret:

Average in-sample regret,

$$\text{Regret}_t = \frac{1}{M_t} \sum_{i,t} \Delta^{D_{it} X_{it}}$$

converges in probability to

$$\gamma \cdot \frac{1}{k} \sum_{x,d} \Delta^{dx} \cdot p^x.$$

4. Converging estimator:

The normalized average outcome for treatment d in stratum x ,

$$\sqrt{M_t} \left(\bar{Y}_t^{dx} - \theta_0^{dx} \right),$$

¹⁵ Note that this statement refers to frequentist consistency (given θ_0) of a Bayesian posterior probability (which averages over θ).

converges in distribution to

$$N\left(0, \frac{\theta_0^{dx}(1 - \theta_0^{dx})}{\bar{q}^{dx} \cdot p^x}\right).$$

The large sample result of Theorem 1 characterizes the trade-offs in choosing γ . The parameter γ allows us to interpolate between non-adaptive, conventional randomization ($\gamma = 1$) and Thompson sampling ($\gamma = 0$). The former is optimal for minimizing the expected variance of treatment effect estimators. The latter is optimal for minimizing the expected regret (maximizing expected welfare) for the participants in the experiment.

As we increase γ , starting from a value of 0, the expected in-sample regret increases linearly in proportion to γ . On the other hand, the asymptotic variance of conditional average treatment effect estimators, comparing the conditionally optimal treatment to its alternatives, is given by one over the total sample size, times

$$\frac{\theta_0^{d^{*x}}(1 - \theta_0^{d^{*x}})}{((1 - \gamma) + \gamma/k) \cdot p^x} + \frac{\theta_0^{dx}(1 - \theta_0^{dx})}{(\gamma/k) \cdot p^x}.$$

This number is decreasing in γ , since higher γ means a more balanced distribution of observations across treatment arms. In our application, we trade off these conflicting objectives by setting the share of observations for which treatment is fully randomized to $\gamma = 0.2$, which implies that the probability of being assigned to each treatment is bounded below by 0.05.

3.4 Discussion of Theorem 1 and the Tempered Thompson Algorithm

Several observations are worth making about the properties of the Tempered Thompson Algorithm and Theorem 1. First, the theorem implies that the large sample properties of the Tempered Thompson Algorithm do not depend on the prior (as long as the latter has full support): In large samples the data dominate the prior, the posterior is consistent, and thus assignment shares become independent of the prior. Relative to pure Thompson sampling, this happens even faster for the Tempered Thompson Algorithm with $\gamma > 0$.

The flip side of this large-sample robustness to the prior is robustness to the data in the

initial periods, for three distinct reasons. First, Bayesian inference optimally combines data and prior, and therefore down-weights outliers among the initial observations. This stabilizes assignment shares in initial periods, and makes them closer to an equal division among treatment arms. Only when evidence has accumulated that some treatment arms are better than others do assignment shares become unequal. Second, relative to Thompson sampling the Tempered Thompson Algorithm additionally shrinks assignment shares toward the balanced assignment. And third, the outcomes in our setting are bounded, and therefore the influence of any single observation on the posterior is necessarily bounded, as well.

The properties of the Tempered Thompson Algorithm listed in Theorem 1 rely on having sufficient power (enough observations) to be able to distinguish the best treatment. It therefore presupposes that the learning environment is sufficiently fast, i.e., the designer can obtain quick feedback. An alternative characterization might consider the worst case behavior of the Tempered Thompson Algorithm across possible values for the parameter vector θ_0 , for a given T . As it turns out, the worst case behavior in terms of in-sample regret is driven by parameter values which are such that the treatment effects Δ^{dx} (relative to the optimal treatment) are of the same order of magnitude as the standard errors for estimates of these treatment effects, that is, of order $1/\sqrt{T}$ (Bubeck and Cesa-Bianchi, 2012; Russo and Van Roy, 2016). If treatment effects are larger, the best treatment is discovered quickly and in-sample regret is low. If treatment effects are smaller, it doesn't matter as much which treatment participants are assigned.¹⁶

3.5 Inference

One worry about adaptive experimental designs is that they lead to biased inference (see for instance Hadad et al. 2019). Item 4 of Theorem 1 implies, however, that this is not the case for the Tempered Thompson Algorithm in large samples. Sample averages in each treatment arm are consistent, asymptotically unbiased, and normally distributed, so that inference can proceed as if treatment assignment was not adaptive. This is true because assignment probabilities for each arm in each stratum are bounded away from 0 when $\gamma > 0$. In our empirical analysis, we nevertheless consider two methods for inference that do not

¹⁶ A formal characterization of the worst-case behavior of the Tempered Thompson Algorithm is left for future work.

rely on such asymptotics, but instead are exactly valid in finite samples despite adaptive assignment, as detailed next.

Our primary form of inference is Bayesian, based on the hierarchical default prior described in Section 3.1 above. To construct credible sets (i.e., sets that have a given posterior probability of containing the true parameters), we report 0.025 and 0.975 quantiles, based on Markov Chain Monte Carlo draws. We do so for all our estimates listed in the previous section. This yields sets that have a posterior probability of 95% to contain the true parameters, conditional on the data of the experiment.

We would like to emphasize that standard Bayesian inference remains valid in finite samples for adaptive designs such as ours, since the likelihood function is not affected by adaptivity. In large samples, as long as $\gamma > 0$, our credible sets also have 95% frequentist coverage probability, i.e., they are confidence sets in the usual sense; cf. [van der Vaart \(2000\)](#), chapter 10. This holds because the share of observations assigned to each treatment in each stratum is bounded below, asymptotically.

Additionally, we provide randomization-based p-values that are valid under the sharp null hypothesis that there are no treatment effects, i.e., under the null that $\theta^{dx} = \theta^{d'x}$ for all d, d', x . Under this null, we can generate counterfactual data by re-running our assignment algorithm repeatedly, leaving outcomes as they are in our data, but generating new treatment assignments. The distribution of test-statistics over this re-randomization distribution can be used to construct critical values and p-values that are exact in finite samples, under the sharp null.

4 Results

In this section, we discuss the impact of the interventions and the performance of Tempered Thompson Algorithm. We first present a set of results on wage employment based on the survey carried out six weeks after baseline. We do this in two different ways. First, we present Bayesian posteriors and credible sets. Second, we report the difference between

weighted average employment in each treatment group and in the control group.¹⁷ Here, we use randomization inference to construct a p -value of the sharp null of no treatment effect. We also present three ‘welfare contrasts’ that quantify the overall impact of our interventions, as detailed in Section 3 below.

We then present a broader set of results from the longer surveys we carried out 2 months and 4 months after baseline. These surveys capture wage employment, but also measure job-search, earnings, well-being, social integration and migration intentions. For each outcome, we report weighted averages and randomization inference p -values as explained above.

4.1 Employment after six weeks and performance of the Tempered Thompson Algorithm

4.1.1 Employment after six weeks

Job-finding rates in the control group are consistently low, especially for Syrians. Six weeks after joining the program, the average control wage-employment rate is 4.9 percent (Table 2). Further, individuals sampled at different points in time tend to have similar six-week employment rates, except for somewhat higher rates for those sampled in the first month of the experiment. We show this in Figure 1 where we plot the employment rate against the week of sampling. These averages, however, mask substantial heterogeneity (Table A.5). Employment rates among Jordanians (6.8 percent) are more than twice as large as employment rates among Syrians (2.7 percent). Similarly, the male employment rate (7.7 percent) is more than twice as large as the female employment rate (3.1 percent). Over-

¹⁷ Weighting is necessary as the samples in each experimental group are mechanically unbalanced due to our adaptive randomization procedure. We report *weighted averages* of the form:

$$\beta_j^d = \frac{1}{N} \sum_{it} \frac{\mathbf{1}(D_{it} = d)}{q^{dx}} \cdot W_{it}^j,$$

where

$$q^{dx} = \frac{\sum_{it} \mathbf{1}(D_{it} = d, X_{it} = x)}{\sum_{it} \mathbf{1}(X_{it} = x)}.$$

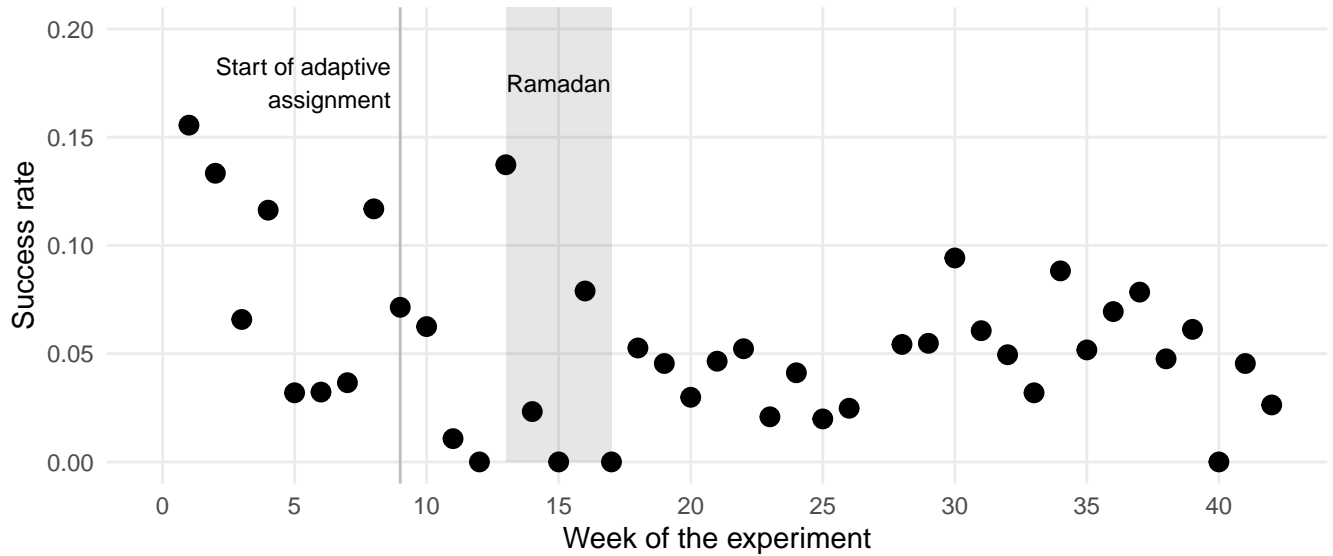
W_{it} is the six weeks employment status of individual i sampled on day t , D_{it} is the treatment status of this individual, X_{it} is the stratum, and N is the total number of experimental participants.

Table 2: **Weighted mean differences in employment after 6 weeks, with randomisation inference p -values**

Treatment	Success rate	Δ	P-value
Cash		0.006	0.296
Information		-0.005	0.690
Nudge		0.003	0.388
Control	0.049		

Note: The table reports results for wage employment at the time of the six weeks follow-up interview. Δ is the difference between weighted mean employment in a given treatment group and in the control group. p -values were obtained with the randomization inference procedure discussed in Section 3.5.

Figure 1: **Employment rate by week of sampling**



all, most subgroups have employment rates below 10 percent.¹⁸ Given that job search at baseline was substantial, this highlights the difficulty of finding work in this labour market.

¹⁸ In Table A.4 we look at the full break-down in sixteen strata, we find that three strata have employment rates above 10 percent. However, in two of these case, the strata have very few observations and so our measure of employment rate is likely to be noisy.

Our main finding is that, six weeks after the start of the program, none of the interventions increase employment for the average program participant. We report Bayesian posteriors on the impacts of the different treatments and the respective credible sets in Figure A.4. These posteriors indicate that the impact on employment is always smaller than 1 percentage point. We confirm this result by reporting differences in weighted employment rates in Table 2.

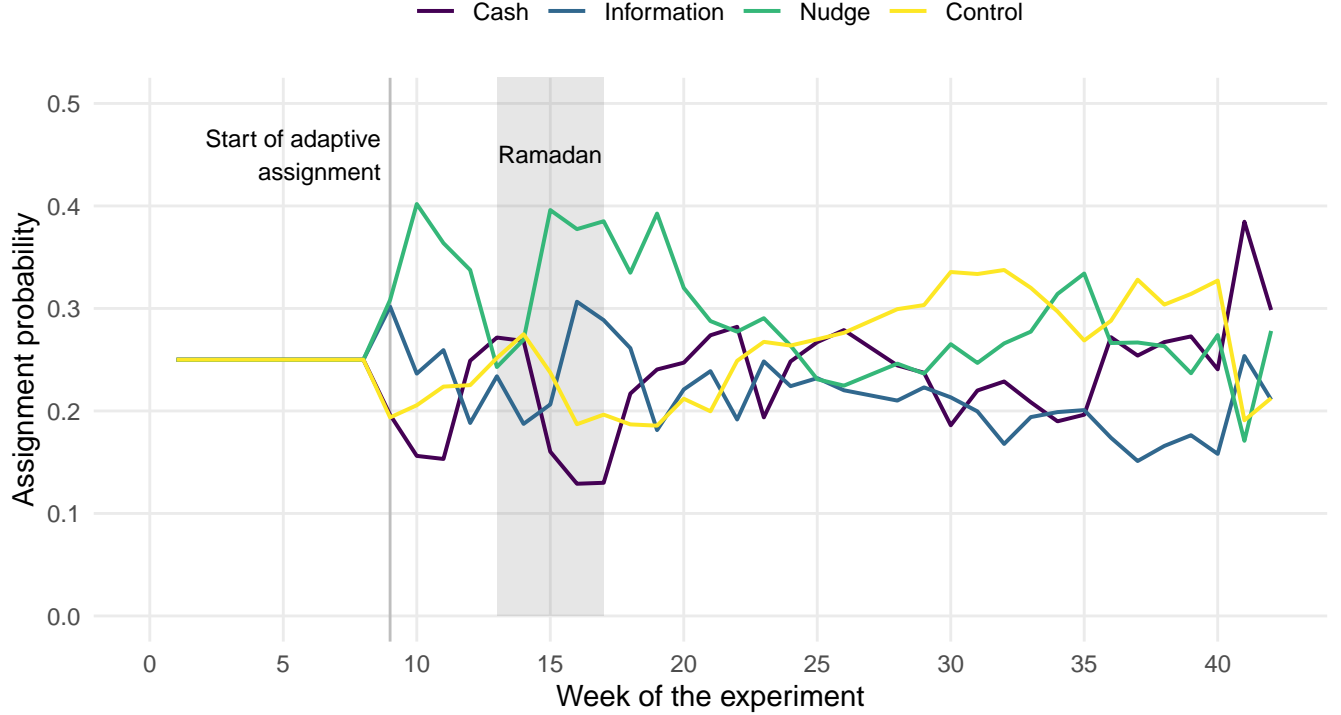
We are unable to find evidence of treatment impacts for specific, pre-specified groups of individuals. In Figure A.4, for example, we show treatment effects after splitting the sample by nationality and do not find evidence of impacts on employment on either Syrians or Jordanians. Posteriors are somewhat larger for Syrians than for Jordanians, but the credible sets always overlap. We report credible sets for all sixteen strata in Table A.6. Further, Table A.5 reports differences in weighted employment by group which confirm these findings. Employment effects are somewhat larger for Syrians (e.g. employment rates in the cash group are a 1.3 percentage point higher than in the control group), but these effects are not significantly different from zero ($p=0.123$).

4.1.2 Performance of the Tempered Thompson Algorithm

Consistently with the results presented above, we find that in the last week of the study our algorithm places similar proportions of people in each of the four experimental groups. We show the probability of assignment to the four experimental conditions for each week of the study in Figure 2. By design, individuals are assigned to the different groups in equal proportion up to the sixth week of the study, as we have no information to update the priors up to that point. When learning started, the algorithm initially assigned more weight to the nudge intervention. However, this was slowly reversed after the 20th week of the study.

The algorithm's departure from equal-proportions randomisation is somewhat more pronounced for specific strata. We show this in Figure 3, where show strata-specific weekly treatment assignment probabilities, and in Table A.7, where we show, for each treatment, the posterior probability that employment rates are highest under that treatment — that is,

Figure 2: Assignment probabilities by week

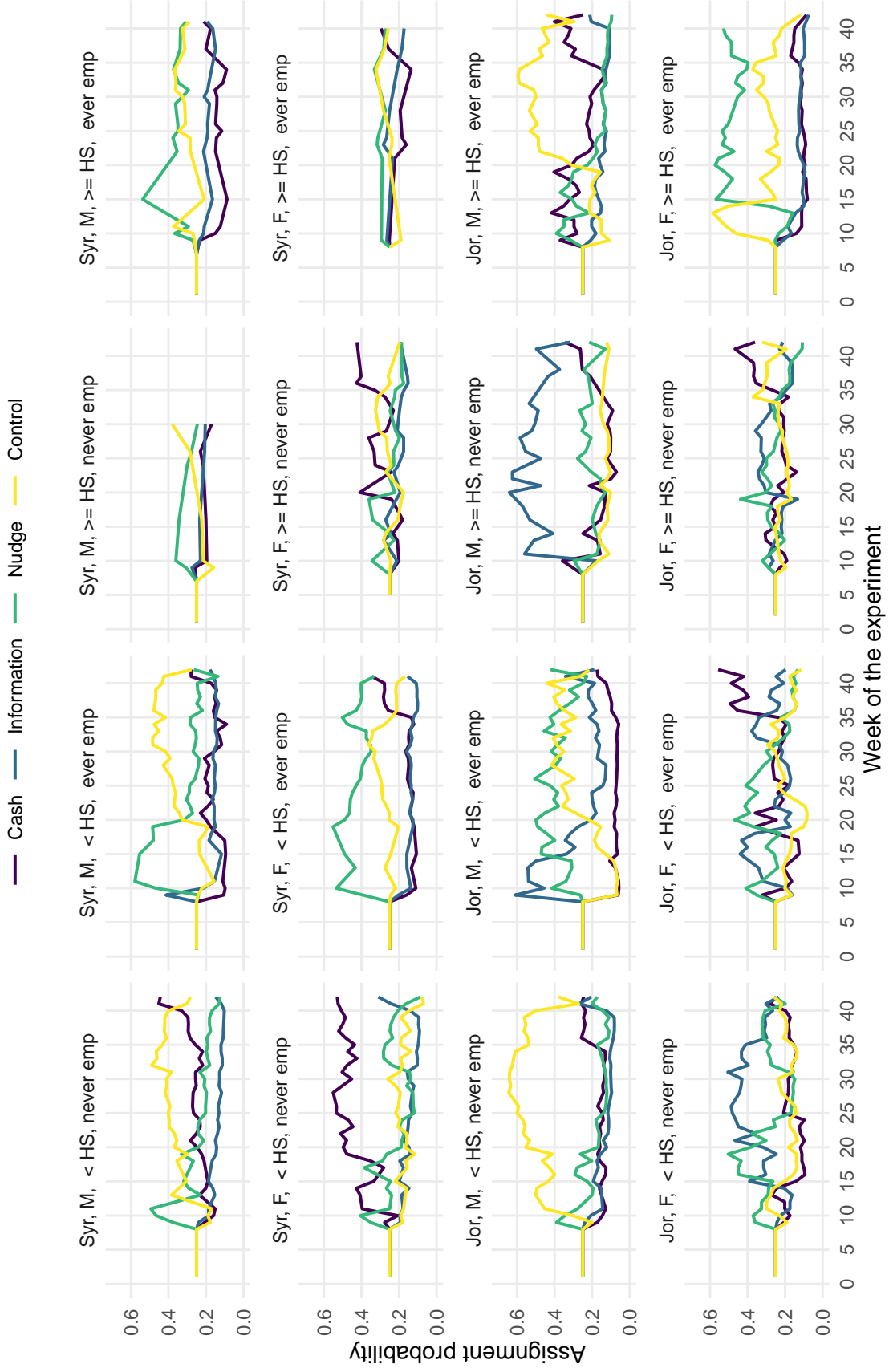


the posteriors that determine treatment assignment probabilities in our algorithm. While for some strata the assignment probabilities never depart from 25% in a sustained way, in some strata we do observe clear changes. For example, in the last week of the experiment, we assign almost 60% of inexperienced and less educated Jordanian women to the cash intervention. Similarly, for some strata, the probability that the control is optimal drops to a few percentage points (e.g. inexperienced, less educated female Syrians). However, it should be stressed that, as discussed above, the differences in potential outcomes we estimate are small and hence the impacts of departing from equal-proportions randomization are limited in this context.

4.1.3 Welfare contrasts

We present three ‘welfare contrasts’ that quantify the overall impact of our interventions, both against a counterfactual where no treatment is given, and against a counterfactual where treatments are randomized in equal proportion. First, within the experiment, we

Figure 3: Assignment probabilities by stratum and by week



compare the average potential outcomes for the *actually chosen treatment* assignment to the average that would have obtained under *random assignment*,

$$\Delta_1 = \frac{1}{N} \sum_{i,t} \left(\hat{E} \left[\theta^{D_{it} X_{it}} \right] - \frac{1}{4} \sum_d \hat{E} \left[\theta^{d X_{it}} \right] \right).$$

This estimate measures how much better we did for our experimental participants, compared to a conventional design with fully random assignment.

Second, we compare the *optimal targeted policy*, and the *optimal non-targeted policy*, to the default of *no intervention* (treatment 0),

$$\begin{aligned} \Delta_2 &= \sum_x \left(\max_d \hat{E} \left[\theta^{dx} \right] - \hat{E} \left[\theta^{0x} \right] \right) p^x, \\ \Delta_3 &= \max_d \sum_x \left(\hat{E} \left[\theta^{dx} \right] - \hat{E} \left[\theta^{0x} \right] \right) p^x. \end{aligned}$$

The definition of Δ_2 allows the optimized d to depend on x , while the definition of Δ_3 requires the same d to be implemented for all x .

We estimate that overall impacts on employment after six weeks are small; Table 3 reports our corresponding estimates of the three welfare contrasts specified above. We have two key findings. First, if we compare the optimal targeted policy to a counterfactual where no intervention is given (welfare contrast Δ_2), we estimate a gain in employment of 1.7 percentage points (95% credible set: [0.001, 0.034]). Relative to the employment rate in the control groups, this amounts to a 35% increase in employment. The optimal non-targeted policy, on the other hand, delivers a gain in employment of about half of a percentage point (welfare contrast Δ_3), with a credible sets that includes zero (95% credible set: [-0.015, 0.27]). The difference in employment gains between these measures suggests that there may be some modest gains from targeting. Overall, the percentage point effects are on the lower end of the impacts of ALMPs on employment reported in McKenzie (2017) (which are typically measured over a longer time frame).

In our study, adaptive randomization did not generate any six-weeks employment gains

Table 3: **Welfare contrasts**

	Estimate	95% Credible set
Δ_1	.002	(0.000,0.004)
Δ_2	.017	(0.001,0.034)
Δ_3	.006	(-0.015,0.027)

Note: The table reports the welfare contrasts defined in Section .

over standard randomization. We show this by reporting welfare contrast Δ_1 , in Table 3, which is very close to zero.

4.2 Longer-term impacts of the interventions

4.2.1 Impacts on job search

Despite the null impacts on employment after six weeks, we document that all interventions generate marked increases in job search among Syrians.¹⁹ As shown in Table 4, the cash transfer raises the proportion of Syrians who look for work two months after baseline by 5.6 percentage points (a 13 percent increase over a control job-search rate of 43 percent) and leads Syrians to place 0.5 more job applications (a 40 percent increase over a control mean of 1.2 applications). Similarly, the information intervention and the nudge intervention raise job search rates by 4.6 percentage points ($p = 0.10$) and 4.2 percentage points ($p = 0.165$) respectively. Both of these interventions also have significant impacts on job applications: a 35 percent increase for the information intervention and a 55 percent increase for the nudge intervention.

Among Jordanians, the cash and information interventions have smaller and insignificant impacts on job search (Table 4). For example, the cash intervention is associated with a 3.2 percentage point, insignificant increase in the job-search rate. However, we document

¹⁹ The analysis of job search outcomes in this sub-section was not pre-specified. In the Pre-Analysis Plan we committed to studying the impacts of the interventions on five main longer-term outcomes (which we report in the next subsection). Further, we anticipated that, motivated by our main results on those outcomes, we would run a number of additional exploratory regressions to better understand treatment mechanisms. This subsection presents this exploratory analysis.

Table 4: Job search impacts after 2 months

	Searched for work	Applications	Hours job search
<i>Syrians</i>			
Cash	0.056 (0.077)	0.518 (0.043)	0.794 (0.133)
Information	0.047 (0.123)	0.423 (0.072)	0.056 (0.482)
Nudge	0.037 (0.195)	0.648 (0.016)	0.698 (0.157)
Control mean	0.437	1.19	4.162
Observations	1536	1440	1444
<i>Jordanians</i>			
Cash	0.033 (0.165)	-0.055 (0.553)	-0.672 (0.831)
Information	0.025 (0.255)	-0.501 (0.874)	-0.457 (0.757)
Nudge	0.065 (0.030)	0.458 (0.130)	0.241 (0.350)
Control mean	0.577	2.71	5.792
Observations	1897	1783	1717

Note: This table reports treatment effects on the three variables capturing job search, 2 months after the baseline interview. ‘Job search’ is a dummy for whether the person has done any job search in the last 30 days. ‘Applications?’ is the number of job applications completed in the last 30 days. ‘Hours search’ is the number of hours of job search in the last seven days. The first panel reports impacts for Syrians, and the second panel reports impacts for Jordanians. Next to each treatment effect, we report a randomization inference *p*-value.

a large and positive impact of the nudge intervention on job search. Importantly, no intervention is associated with a significant impact on job applications among Jordanians and, for both the cash and information intervention, the effect is actually negative. Finally, Table 4 also highlights that Jordanians search much more intensely than Syrian refugees in the absence of the interventions: the control job search rate is 30 percent higher and the control number of job applications is twice as large.

4.2.2 Impacts on labour market outcomes

We find that the both the cash and information intervention improve Syrian refugees' labour market outcomes, up to 4 months after baseline.²⁰ We report the relevant coefficient estimates and randomisation inference p -values in Table 5 and Table 6. Offering cash leads to a significant increase in the employment rate of more than 50 percent (an effect of 4.8 percentage points in the 2 month survey and 3.8 percentage points in the 4 month survey) and a significant boost in earnings of about 40 percent after 2 months and of 65 percent after 4 months. The information intervention increased employment and earnings by almost the same amount as the cash grant two months after baseline. Four month after baseline, this intervention generates a 40 percent increase in employment (marginally insignificant) and a significant 55 percent increase in earnings. We also find that the nudge intervention has weaker and short-lived effects on the labour market outcomes of refugees. Four months after baseline, we are unable to document any significant impacts of this intervention.²¹

The magnitude of the effects on refugee employment and earnings that we document is large relative to the estimates reported in the recent literature for other active labor market policies in developing countries (McKenzie, 2017). In proportional terms, both the earning and the employment effects are at the top of the distribution of the estimates reported in the literature. While this is partly driven by the low control employment rate, in absolute terms, the employment effect is still close to the top of the distribution of existing estimates for job search assistance policies (but smaller than the impacts of the most effective wage

²⁰ The analysis in this sub-section was pre-specified. We summarise all variable definitions in Table A.8. We report results disaggregated by nationality in Table 5 and 6. We report results for the full sample in Table A.10 and A.11 in the Appendix.

²¹ All employment impacts reported in the paper refer to 'wage employment'. Due to ethical and confidentiality concerns, we did not ask refugees to specify whether a particular job was formal or not. Individuals may have been particularly reluctant to share information about informal employment during the 6 week follow-up call, as this interview was carried by a single enumerator who had not interacted with respondents before. Conversely, the two and four-month interviews were carried out by the same person who had interviewed the respondent at baseline and had enrolled them in the program. Respondents, in particular Syrian refugees, may have had more trust in this enumerator and hence may have been more likely to discuss informal work with them. This may partly explain the difference in the control employment rate and in treatment effects that we observe when we compare the 6 week and the 2 month interviews. In support of this interpretation, we show in Table A.9 that a substantial share of the increase in wage employment documented in the 2 month follow-up survey occurs in jobs that pay just below the formal minimum wage, a proxy of informality.

subsidy and training interventions).²² In terms of cost-effectiveness, the cash intervention would pay for itself through higher earnings if the 4 month impacts were sustained for about one year.

For refugees, the cash grant and information interventions are also associated with small, insignificant increases in the well-being index and with an insignificant 7 percent drop in the proportion of people that intend to migrate outside of Jordan. While the migration effect is not significant, in absolute terms, the 4 percentage points decrease we estimate is commensurate to the size of the positive employment effects of these interventions. Finally, we do not document any impacts on social integration, consistent with recent evidence suggesting that baseline social integration for Syrian refugees in Jordan is high relative to the experience of Syrian refugees in European countries or in the US ([Alrababa'h et al., 2019](#)).

For Jordanians, on the other hand, we are unable to find evidence of labour market impacts for any intervention (Table 5 and Table 6). This is particularly surprising for the nudge intervention, which has positive impacts on job search for this population. Further, for the cash intervention, we document an (insignificant) 2.9 percentage point reduction in employment and an (insignificant) 25 percent reduction earnings after 4 month, but a contemporaneous, significant increase in the well-being index of 0.06 of a standard deviation. This may be consistent with the cash intervention enabling jobseekers to reject offers for undesirable jobs.

4.2.3 What prevents effective job search among refugees?

What have we learned about barriers to job search among refugees? Our experimental estimates show that the cash intervention has the largest impacts on job search and employment for this population. Here, we present evidence suggesting that the cash grant is effective because liquidity constraints are a key labour market barrier for refugees.²³ We do this by studying whether proxies for available liquidity are associated with higher control

²² It should be noted that the effects we report in this paper are on a shorter time-frame than most of the estimates reported in [McKenzie \(2017\)](#).

²³ The analysis in this section was not pre-registered, but is part of our exploration of treatment mechanisms.

job search intensity and whether treatment effects are heterogeneous with respect to liquidity. Further, we report impacts on additional measures of job quality. Liquidity-constrained individuals would forgo desirable employment opportunities due to the inability to pay for search and application costs; if these constraints are binding, the marginal jobs obtained by cash beneficiaries would have similar or better quality as the control jobs.

First, among control refugees, we find a strong association between job-search intensity and expenditure at baseline — a proxy of liquidity. We plot this relationship non-parametrically in Figure A.6 in the appendix; both the probability of searching for work and the number of job applications increase with expenditure, especially for individuals with expenditure below the median. These associations are sizeable. Using a linear regression, we find that a one standard deviation increase in expenditure at baseline is associated with a 0.6 standard deviation increase in the number of job applications sent (and a 0.08 standard deviation increase in the probability of job search). In contrast, among Jordanians, this relationship is much weaker: an increase in expenditure by one standard deviation is associated with a 0.3 standard deviation increase in the number of job applications and a 0.03 standard deviation increase in the probability of job search (also see Figure A.7).

Second, we show that the impacts of the cash intervention are driven by refugees who have expenditure below the median. We show this in Figure A.8, A.9 and A.10: impacts on job search and employment are concentrated among the poorest respondents (while impacts on job applications are more evenly distributed). In contrast, the information and nudge intervention have (i) generally weaker impacts on job search among low-expenditure refugees, and (ii) employment impacts for low expenditure refugees that are about half of those of the cash intervention. Additional evidence in support of credit constraints comes from refugees' reports on how they spent the cash: 26 percent of recipients in the low-expenditure group report that they mostly spend the money on job search. Among above-median expenditure recipients, this proportion drops to 18 percent.²⁴

Third, we find that the cash intervention boosts job retention and, after four months, hourly

²⁴ Among Jordanians, 32 percent of respondents report to have spent the cash mostly on job search. However, this proportion does not vary by baseline expenditure. Given the null impacts on job search and the higher control job-search intensity, it is likely that cash given to Jordanians has mostly financed infra-marginal job search.

wages; we show this in Table 7. The grant doubles the probability of having retained a job between the two and the four-month interview — from 3.3 percent to 6.2 percent. Further, mean hourly wages among employed cash beneficiaries are .63 of a standard deviation higher than in the control group. The other two interventions, on the other hand, are associated with much smaller increases in retention and hourly wages (for example, job retention among information recipients is 5.2 percent, and the impact on hourly wages is about .04 of a standard deviation). These impacts indicate that, consistently with the prediction of a model where job search is constrained by limited liquidity, the cash intervention enables jobseekers to find jobs that have higher match quality — and hence are more stable and better paid.

Table 5: Treatment effects on main outcomes after 2 months

	Employed	Earnings	Well-being	Social integration	Intends to migrate
<i>Syrians</i>					
Cash	0.052 (0.017)	7.204 (0.062)	0.021 (0.333)	-0.009 (0.575)	-0.038 (0.838)
Information	0.047 (0.036)	6.209 (0.092)	0.025 (0.309)	-0.035 (0.728)	-0.041 (0.856)
Nudge	0.035 (0.081)	4.210 (0.185)	0.007 (0.445)	-0.055 (0.817)	-0.031 (0.783)
Control mean	0.091	16.268	0.088	0.011	0.664
Observations	1608	1605	1608	1608	1598
<i>Jordanians</i>					
Cash	-0.007 (0.618)	-1.491 (0.618)	0.101 (0.012)		
Information	-0.006 (0.624)	-2.486 (0.696)	0.019 (0.342)		
Nudge	-0.004 (0.585)	-1.684 (0.631)	0.015 (0.385)		
Control mean	0.128	29.22	0.069		
Observations	1985	1977	1985		

Note: This table reports treatment effects on the five main outcomes specified in the Pre-Analysis Plan, 2 months after the baseline interview. ‘Employed’ is a dummy for whether the person has a wage-paying job at the time of the interview. ‘Earnings’ is the value earnings from the main job (where individuals who are not in wage employment are assigned a zero). ‘Well-being’ is a weighted index that includes: (i) a measure of expenditure, (ii) a measure of positive affect, and (iii) a measure of life satisfaction. ‘Social integration’ is an index of social integration. ‘Intends to migrate’ is a dummy for whether the respondent intends to migrate to a third country (i.e. this measure does not include return migration). The first panel reports impacts for Syrians, and the second panel reports impacts for Jordanians. Next to each treatment effect, we report a randomization inference *p*-value.

Table 6: Treatment effects on main outcomes after 4 months

	Employed	Earnings	Well-being	Social integration	Intends to migrate
<i>Syrians</i>					
Cash	0.038 (0.027)	6.550 (0.040)	0.043 (0.163)	0.005 (0.472)	-0.044 (0.875)
Information	0.019 (0.148)	4.567 (0.105)	0.003 (0.480)	0.001 (0.470)	-0.046 (0.884)
Nudge	0.003 (0.449)	0.260 (0.484)	0.052 (0.106)	0.005 (0.467)	-0.034 (0.810)
Control mean	0.052	9.76	0.008	-0.005	0.675
Observations	1565	1563	1565	1565	1561
<i>Jordanians</i>					
Cash	-0.025 (0.855)	-7.515 (0.911)	0.068 (0.055)		
Information	-0.009 (0.658)	-4.299 (0.773)	0.041 (0.169)		
Nudge	-0.002 (0.544)	-2.845 (0.709)	0.040 (0.176)		
Control mean	0.144	33.451	0.039		
Observations	1913	1900	1913		

Note: This table reports treatment effects on the five main outcomes specified in the Pre-Analysis Plan, 2 months after the baseline interview. ‘Employed’ is a dummy for whether the person has a wage-paying job at the time of the interview. ‘Earnings’ is the value earnings from the main job (where individuals who are not in wage employment are assigned a zero). ‘Well-being’ is a weighted index that includes: (i) a measure of expenditure, (ii) a measure of positive affect, and (iii) a measure of life satisfaction. ‘Social integration’ is an index of social integration. ‘Intends to migrate’ is a dummy for whether the respondent intends to migrate to a third country (i.e. this measure does not include return migration). The first panel reports impacts for Syrians, and the second panel reports impacts for Jordanians. Next to each treatment effect, we report a randomization inference *p*-value.

Table 7: Retention and wages for Syrians

	Job retention month 4	Hourly wage month 2	Hourly wage month 4
Cash	0.030 (0.036)	0.037 (0.393)	0.308 (0.055)
Information	0.017 (0.139)	-0.131 (0.835)	-0.087 (0.673)
Nudge	0.011 (0.270)	-0.124 (0.812)	-0.010 (0.527)
Control mean	0.034	1.377	1.197
Observations	1565	193	94

4.3 Eliciting expert forecasts

At the time of launching our interventions, we conducted an incentivized elicitation exercise with IRC staff. We ran this online, surveying 16 staff based in Amman, and four senior staff based in New York. For all respondents, we began by providing descriptive quantitative information on the background of our sample, a brief description of each intervention, and information on the employment rate for a similar sample in 2018 (namely, a rate of 2.5%). We then asked a series of questions about each respondent's prediction for the rate of employment after six weeks; this was directly incentivized.²⁵ We illustrate forecast employment rates in the Online Appendix (Figures A.11 and A.12).

We have three main findings. First, relative to our estimated treatment effects, local staff were very optimistic in their forecasts: at the median, they predicted employment rates of 20%, 10% and 9% for the cash, information and nudge interventions respectively (the median forecast for the control group was 2.25%). Senior staff had more accurate forecasts: medians of 7%, 5% and 4% (against a median prediction in the control group of 3%). Second, the dispersion in forecasts was very large, indicating substantial uncertainty about treatment impacts. Third, interestingly, both local and senior staff correctly anticipated that the cash intervention would be most effective.

4.4 Qualitative fieldwork

Five months after the trial began, we conducted structured qualitative interviews in the form of focus group discussions with participants who have received one of the three search interventions. The purpose of the interviews was to build a deeper understanding of the job search process and the mechanisms by which some interventions may have worked for different groups of participants. Participants were divided into six single-gender groups, each group focusing on one of the three interventions (cash, informational, or psychological support). We found the following results.

First, consistently with our experimental findings, respondents identified financial con-

²⁵ Specifically, we told respondents that we would randomly draw one of their employment forecasts; if this forecast was within 1 percentage point of the correct answer, we would provide a lottery ticket having 50 tickets and a prize of US\$200.

straints as a key barrier to economic opportunity. The cash grant was by far the most popular intervention, irrespective of the demographics of the respondents. Among those who received the grant, some indicated that they used the cash directly to cover transportation costs when searching for work. Several others, in particular, Syrian females, reported using the cash to cover immediate basic needs such as medical bills for themselves or the family. Given participants' highly vulnerable economic situations, the cash was seen as necessary step before searching for work. One Syrian female cash recipient reported that she used the cash for medical care, which allowed her to then begin searching for work and eventually to find a job in a factory. These qualitative findings match our quantitative results.

Second, participants reported mixed views on the informational and nudge interventions. The information intervention was only reported to be useful or accessible by Jordanians who had higher education levels than the typical participant in our study. At the same time, for some of those who understood the content, the information intervention was a source of frustration given that employers frequently violated labor laws. As such, the information on labor laws was not deemed to represent *de facto* labor rights.

The nudge intervention was reported to be useful by the Syrian women we talked to, but for reasons quite different from our original theory of change. Rather than working as a commitment device, these respondents reported that the intervention motivated them by making them feel 'like someone cared about them', and by making job search top of mind. One Syrian woman indicated that following the last SMS message, she continued to use her tool and shared it with her female neighbours. However, others (mostly Jordanian males) reported a lack of interest or desire to receive SMS reminders on job search intentions and achievements. For this group of participants, the intervention was perceived not to be useful given they were already searching for work.

Finally, participants in all focus groups stressed that formal jobs are often poorly paid and offer bad working conditions. Women reported concerns about working in the formal manufacturing sector, where they would be required to mix with men and undergo long hours away from their children. In fact, many women indicated a strong interest in starting home-based businesses like a kitchen, beauty salon, accessory shop, or nursery which would offer

them both a ‘safer’ and more flexible work environment. Men also perceived formal work in factories to be undesirable, but for different reasons. They mostly complained about low pay: wages in manufacturing firms are considered insufficient to support a family. Per diem informal work, mainly in services like plumbing, carpentry, and painting, was seen as more profitable. Additionally, factory work was seen as being potentially exploitative and men shared stories about wages being withheld or delayed by employers for arbitrary reasons. These observations suggest that constraints to labor market participation among both refugees and Jordanians are by no means restricted to the labor market frictions we have focused on in this study.

5 Conclusion

Randomized controlled trials have come under criticism from an ethical perspective. For example, [Deaton \(2020, p. 21\)](#) points out that “It is particularly worrying if the research addresses questions in economics that appear to have no potential benefit for the subjects.” Relatedly, implementation partners might be reluctant to engage in continued experimentation if they believe that they already know which intervention works best. Adaptive experimentation can help mitigate both ethical criticisms of RCTs and the reluctance of implementation partners to engage in experimentation, by setting welfare maximization (or regret minimization) as the main objective of an experiment. Indeed, any optimal adaptive experimental design has the property that participant welfare cannot be increased in the long-run by using any other (experimental or non-experimental) treatment allocation procedure.

In this paper, we have reported the results of an implementation of adaptive targeted treatment allocation in a field experiment. Our Tempered Thompson Algorithm strikes a balance between maximizing participant welfare and providing precise estimates of treatment effects. Our implementation context was novel: We looked at the effects of active labor market policies on Syrian refugees and local job-seekers in Jordan. Our treatments did not appear to have a significant effect on refugee employment after six weeks, but some of the interventions appear to have had a substantial impact on longer-term employment outcomes.

Our results show that adaptive targeted experiments can be straightforwardly deployed in the field and can be used to draw scientific and policy conclusions. Moreover, our methodology creates many possibilities for further applications. The Tempered Thompson Algorithm is a powerful tool for any setting in which subjects arrive over time and their outcomes are observed within a short time-frame. In addition to employment programs, our methodology may be applied in many other development contexts, including drug and vaccination programs, agricultural technology adoption programs, and emergency relief programs. Further work on adaptive experiments can tailor experimental designs to specific applications.

There remains, however, the thorny issue of how to actually define and measure participant welfare. In many contexts such as employment, education, or health, the experimentalist would like to target long-term participant outcomes. However, adapting treatment allocation based on long-term outcomes can make the field experiment too long and costly. Instead of only measuring and targeting long-term outcomes, the designer might therefore wish to find a set of short-run proxies, i.e., “statistical surrogates”, for long-term welfare ([Athey et al., 2019](#)). An adaptive targeted field experiment would therefore be designed in order to target these statistical surrogates. In our case, the choice of target outcome, i.e., formal employment 6 weeks after the intervention, was mainly driven by the organizational objectives of our implementation partner and their donors. Our results (Section 4) suggest that if we had targeted informal employment and if informal employment were an appropriate proxy for long-term welfare, our experiment would have substantially increased participant welfare.²⁶

²⁶ The challenge of aligning decisions based on machine learning (e.g., of robots and other artificial intelligence systems) with broader societal interests is by no means unique to adaptive experiments ([Taylor et al., 2016](#)).

References

- Abebe, G. T., S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2020). Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City.
- Abel, M., R. Burger, E. Carranza, and P. Piraino (2019). Bridging the Intention-behavior Gap? The Effect of Plan-making Prompts on Job Search and Employment. *American Economic Journal: Applied Economics* 11(2), 284–301.
- Agrawal, S. and N. Goyal (2012). Analysis of Thompson sampling for the multi-armed bandit problem. In *Conference on Learning Theory*, pp. 39–1.
- Agrawal, S. and N. Goyal (2013). Further optimal regret bounds for thompson sampling. In *Artificial intelligence and statistics*, pp. 99–107.
- Alfonsi, L., O. Bandiera, V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2020). Tackling Youth Unemployment: Evidence from a Labor Market Experiment in Uganda. *Econometrica*.
- Alrababa'h, A., S. Williamson, A. B. Dillon, D. Hangartner, and J. Hainmueller (2019). Attitudes toward migrants in a highly impacted economy: evidence from the syrian refugee crisis in jordan. *Comparative Political Studies*.
- Amjad, R., J. Aslan, E. Borgnäs, D. Chandran, E. Clark, A. Ferreira dos Passos, J. Joo, and O. Mohajer (2017, July). Examining Barriers to Workforce Inclusion of Syrian Refugees in Jordan. <https://betterwork.org/wp-content/uploads/2017/08/Formatted-Final-SIPA-Capstone-1.pdf>.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association* 103(484), 1481–1495.
- Athey, S., R. Chetty, G. W. Imbens, and H. Kang (2019, November). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Working Paper 26463, National Bureau of Economic Research.
- Balboni, C., O. Bandiera, R. Burgess, M. Ghatak, and A. Heil (2020). Why do people stay poor?

-
- Banerjee, A. V., R. Hanna, G. E. Kreindler, and B. A. Olken (2017). Debunking the stereotype of the lazy welfare recipient: Evidence from cash transfer programs. *The World Bank Research Observer* 32(2), 155–184.
- Banerjee, A. V. and A. F. Newman (1993). Occupational choice and the process of development. *Journal of Political Economy* 101(2), 274–298.
- Banerjee, A. V. and S. Sequeira (2020). Spatial mismatches and imperfect information in the job search.
- Barbelet, V., J. Hagen-Zanker, and D. Mansour-Ille (2018, February). The Jordan Compact: Lessons learnt and implications for future refugee compacts. <https://www.odi.org/sites/odi.org.uk/files/resource-documents/12058.pdf>.
- Bassi, V. and A. Nansamba (2020). Information Frictions in the Labor Market: Evidence from a Field Experiment in Uganda. *Working Paper*.
- Battisti, M., Y. Giesing, and N. Laurensyeva (2019). Can job search assistance improve the labour market integration of refugees? evidence from a field experiment. *Labour Economics* 61, 101745.
- Benhassine, N., F. Devoto, E. Duflo, P. Dupas, and V. Pouliquen (2015). Turning a Shove into a Nudge? A ‘Labeled Cash Transfer’ for Education. *American Economic Journal: Economic Policy* 7(3), 86–125.
- Bubeck, S. and N. Cesa-Bianchi (2012). Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *Foundations and Trends in Machine Learning* 5(1), 1–122.
- Caliendo, L., M. Dvorkin, and F. Parro (2015). The Impact of Trade on Labor Market Dynamics. Technical report, National Bureau of Economic Research.
- Chamberlain, G. (2011). Bayesian aspects of treatment choice. *The Oxford Handbook of Bayesian Econometrics*, 11–39.
- Deaton, A. (2020, July). Randomization in the tropics revisited: a theme and eleven variations. Technical Report 27600, National Bureau of Economic Research.

-
- Dehejia, R. H. (2005). Program evaluation as a decision problem. *Journal of Econometrics* 125(1-2), 141–173.
- DellaVigna, S. and M. D. Paserman (2005). Job Search and Impatience. *Journal of Labor Economics* 23(3), 527–588.
- Devictor, X. and Q.-T. Do (2017). How Many Years Have Refugees Been in Exile? *Population and Development Review* 43(2), 355–369.
- DSP and Columbia (2020, January). In My Own Hands: A medium-term approach towards self-reliance and resilience of Syrian refugees and host communities in Jordan. <http://dsp-syria.org/sites/default/files/2020-02/DSP-CU%20report.pdf>.
- Duflo, E. and A. Banerjee (2017). *Handbook of field experiments*. Elsevier.
- Fallah, B., C. Krafft, and J. Wahba (2019). The impact of refugees on employment and wages in Jordan. *Journal of Development Economics* 139, 203–216.
- Franklin, S. (2018). Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies. *The Economic Journal* 128(614), 2353–2379.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2014). *Bayesian data analysis*, Volume 2. Taylor & Francis.
- Ghosal, S. and A. Van der Vaart (2017). *Fundamentals of nonparametric Bayesian inference*, Volume 44. Cambridge University Press.
- Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)* 41(2), 148–164.
- Gordon, G. (2017). Solving the Refugee Employment Problem in Jordan: A Survey of Syrian Refugees. *International Rescue Committee report*.
- Government of Jordan (2016, February). The Jordan Compact: A new holistic approach between the Hashemite Kingdom of Jordan and the International Community to deal with the Syrian refugee crisis. <https://reliefweb.int/report/jordan/jordan-compact-new-holistic-approach-between-hashemite-kingdom-jordan-and>.
- Government of Jordan (2019). Jordan Response Plan for the Syrian Crisis: Draft manuscript.

-
- Groh, M., N. Krishnan, D. McKenzie, and T. Vishwanath (2016a). Do wage subsidies provide a stepping-stone to employment for recent college graduates? evidence from a randomized experiment in Jordan. *Review of Economics and Statistics* 98(3), 488–502.
- Groh, M., N. Krishnan, D. McKenzie, and T. Vishwanath (2016b). The impact of soft skills training on female youth employment: evidence from a randomized experiment in Jordan. *IZA Journal of Labor & Development* 5(1), 9.
- Groh, M., D. McKenzie, N. Shammout, and T. Vishwanath (2015). Testing the importance of search frictions and matching through a randomized experiment in Jordan. *IZA Journal of Labor Economics* 4(1), 1–20.
- Groh, M., D. McKenzie, and T. Vishwanath (2015). Reducing information asymmetries in the youth labor market of Jordan with psychometrics and skill based tests. *The World Bank Economic Review* 29(suppl_1), S106–S117.
- Hadad, V., D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey (2019). Confidence intervals for policy evaluation in adaptive experiments.
- Herkenhoff, K., G. Phillips, and E. Cohen-Cole (2016). How Credit Constraints Impact Job Finding Rates, Sorting & Aggregate Output. Technical report, National Bureau of Economic Research.
- IRC (2017, February). In Search of Work: Creating Jobs for Syrian Refugees: A Case Study of the Jordan Compact. <https://www.rescue.org/sites/default/files/document/1343/insearchofworkweb.pdf>.
- Kasy, M. (2018). Optimal taxation and insurance using machine learning—sufficient statistics and beyond. *Journal of Public Economics* 167, 205–219.
- Kasy, M. and A. Sautmann (2019). Adaptive treatment assignment in experiments for policy choice. Technical report. <https://maxkasy.github.io/home/files/papers/adaptiveexperimentpolicy.pdf>.
- Kasy, M. and A. Teytelboym (2020a). Adaptive combinatorial allocation. In preparation.
- Kasy, M. and A. Teytelboym (2020b). Adaptive targeted infectious disease testing. *Oxford Review of Economic Policy*.

-
- Kaufmann, E., N. Korda, and R. Munos (2012). Thompson sampling: An asymptotically optimal finite-time analysis. In *International Conference on Algorithmic Learning Theory*, pp. 199–213. Springer.
- Kitagawa, T. and A. Tetenov (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica* 86(2), 591–616.
- Kluve, J., S. Puerto, D. Robalino, J. M. Romero, F. Rother, J. Stöterau, F. Weidenkaff, and M. Witte (2019). Do youth employment programs improve labor market outcomes? A quantitative review. *World Development* 114, 237–253.
- McKenzie, D. (2017). How Effective are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence. *The World Bank Research Observer* 32(2), 127–154.
- Meager, R. (2019). Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics* 11(1), 57–91.
- Melfi, V. F. and C. Page (2000). Estimation after adaptive allocation. *Journal of Statistical Planning and Inference* 87(2), 353–363.
- Razzaz, S. (2017). A Challenging Market Becomes More Challenging. https://www.ilo.org/wcmsp5/groups/public/---arabstates/---ro-beirut/documents/publication/wcms_556931.pdf.
- Russo, D. (2016). Simple bayesian algorithms for best arm identification. In *Conference on Learning Theory*, pp. 1417–1418.
- Russo, D. and B. Van Roy (2016). An information-theoretic analysis of thompson sampling. *The Journal of Machine Learning Research* 17(1), 2442–2471.
- Russo, D. J., B. Van Roy, A. Kazerouni, I. Osband, Z. Wen, et al. (2018). A tutorial on thompson sampling. *Foundations and Trends in Machine Learning* 11(1), 1–96.
- Scott, S. L. (2010). A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry* 26(6), 639–658.

-
- Shami, S. (2019). When Worlds Collide: Lessons learned from the intersection of behavioral and human-centered design in humanitarian contexts. <https://medium.com/airbel/lessons-learned-from-the-intersection-of-behavioral-and-human-centered-design-in-humanitarian-contexts>.
- Taylor, J., E. Yudkowsky, P. LaVictoire, and A. Critch (2016). Alignment for advanced machine learning systems. Technical report, Machine Intelligence Research Institute.
- Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25(3/4), 285–294.
- UNHCR (2019a). UNHCR: Figures at a Glance. <https://www.unhcr.org/en-us/figures-at-a-glance.html>.
- UNHCR (2019b, November). Updates from Ministry of Labor (MoL) on Syrian refugees’ work permits. Presentation delivered at the November 2019 UNHCR Livelihoods Working Group.
- UNHCR (2020, March). Syria emergency. <https://www.unhcr.org/syria-emergency.html>.
- van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press.
- Verme, P., C. Gigliarano, C. Wieser, K. Hedlund, M. Petzoldt, and M. Santacroce (2015). *The welfare of Syrian refugees: evidence from Jordan and Lebanon*. The World Bank.
- Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* 113(523), 1228–1242.

Appendix

A.1 Proofs

A.1.1 Preliminaries

Our characterization of the large sample properties of our γ -Thompson algorithm relies on the following two useful results from the literature. The first is a law of large numbers for adaptive sequences, which can be found as Lemma 5 in [Russo \(2016\)](#). The second is a sufficient condition for consistency of Bayesian posteriors, known as Schwartz's theorem, which can be found as Theorem 6.16 in [Ghosal and Van der Vaart \(2017\)](#).

Lemma 1 (LLN for adaptive sequences) *Let $\{Y_n\}$ be an i.i.d sequence of real-valued random variables with finite variance and let $\{W_n\}$ be a sequence of binary random variables. Suppose each sequence is adapted to the filtration $\{\mathcal{H}_n\}$, and define $Z_n = P(W_n = 1 | \mathcal{H}_{n-1})$. If, conditioned on \mathcal{H}_{n-1} , each Y_n is independent of W_n , then with probability 1,*

$$\lim_{n \rightarrow \infty} \sum_{l=1}^n Z_l = \infty \Rightarrow \lim_{n \rightarrow \infty} \frac{\sum_{l=1}^n W_l Y_l}{\sum_{l=1}^n Z_l} = E[Y_1].$$

Theorem 2 (Schwartz) *If $p_0 \in KL(\Pi)$ and for every neighborhood \mathcal{U} of p_0 there exist tests φ_n such that $P_0^n \varphi_n \rightarrow 0$ and $\sup_{p \in \mathcal{U}_c} P^n(1 - \varphi_n) \rightarrow 0$, then the posterior distribution $\Pi(\cdot | X, \dots, X)$ in the model $X, \dots, X | p \sim^{iid} p$ and $p \sim \Pi$ is strongly consistent at p_0*

In the statement of this theorem, Π is the prior distribution, $KL(\Pi)$ is its Kullback-Leibler support.

A.1.2 Proof of Theorem 1

Let $W_{it} = \mathbf{1}(D_{it} = d, X_{it} = x)$, and

$$Z_{it} = E_t[W_{it}] = \left((1 - \gamma) \cdot p_t^{dx} + \gamma/k \right) \cdot p^x,$$

where E_t denotes the conditional expectation given observations up to wave $t - 1$, and conditional on θ . We can rewrite the sample average as

$$\bar{Y}_t^{dx} = \frac{\sum_{i,t' \leq t} W_{it'} Y_{it'}}{\sum_{i,t' \leq t} Z_{it'}} \cdot \frac{\sum_{i,t' \leq t} Z_{it'}}{\sum_{i,t' \leq t} W_{it'}}.$$

We have by construction that $Z_{it} \geq p^x \cdot \gamma/k$, and since $N_t \geq \underline{N}$, it follows that $\sum_{i,t' \leq t} Z_{it'} \rightarrow \infty$ as $t \rightarrow \infty$. Applying Lemma 1 to the first fraction, and a standard law of large numbers to the inverse of the second fraction, we get that

$$\bar{Y}_t^{dx} \rightarrow \theta_0^{dx}$$

in probability as $t \rightarrow \infty$.

1. Given the assumed uniqueness of d^{*x} , there exists an ϵ -neighborhood of θ_0 such that d^{*x} is constant for all x in this neighborhood. The claim follows if we can show that the posterior probability of such an ϵ -neighborhood goes to 1 in probability as $t \rightarrow \infty$.

Given our assumption that the prior for θ has full support, this condition follows from Schwartz's theorem (Theorem 2), if we can show existence of a consistent test for the hypothesis that $\theta = \theta_0$ against the alternative that $\|\theta - \theta_0\| > \epsilon$.

In our setting such a test can be constructed by setting

$$\varphi_t = \mathbf{1}(\|\bar{Y} - \theta_0\| > \epsilon/2).$$

The required consistency follows by convergence in probability of \bar{Y} .

2. By construction of our algorithm, treatment d is assigned with probability $(1 - \gamma) \cdot p_t^{dx} + \gamma/k$ to units in stratum x in period t . It follows from item 1 that this probability converges to \bar{q}_t^{dx} as $t \rightarrow \infty$.

Since N_t is bounded below, the same holds for the cumulative share \bar{q}_t^{dx} .

3. By definition,

$$\text{Regret}_t = \sum_{x,d} \Delta^{dx} \bar{q}_t^{dx} \bar{p}_t^x,$$

where \bar{p}_t^x is the share of observations in stratum x up to period t . The claim follows

from item 2, and the law of large numbers for \bar{p}_t^x , once we note that $\Delta^{dx} = 0$ for $d = d^{*x}$.

4. This is an immediate consequence of Corollary 2.1 and Theorem 3.2 in [Melfi and Page \(2000\)](#), where the necessary conditions of their Theorem 3.2 are verified by our item 2.

■

A.2 Markov Chain Monte Carlo

Algorithm 1 Markov Chain Monte Carlo for the hierarchical Bayes model

Require: The cumulated assignment frequencies m^{dx} and success numbers r^{dx} .

Starting values α_0, β_0 , length of the burn in period B , and number of draws R .

- 1: **for** $\rho = 1$ **to** $B + R$ **do**
- 2: Gibbs step:
 Given $\alpha_{\rho-1}$ and $\beta_{\rho-1}$, for all d, x
 draw θ^{dx} from the $Beta(\alpha_{\rho}^d + r^{dx}, \beta_{\rho}^d + m^{dx} - r^{dx})$ distribution.
- 3: Metropolis step 1:
 Given $\beta_{\rho-1}$ and θ_{ρ} , draw α_{ρ}^d
 by sampling from a normal proposal distribution (truncated below).
 Accept this draw if an independent uniform draw is less than the ratio of the posterior for the new draw, relative to the posterior for $\alpha_{\rho-1}^d$.
 Otherwise set $\alpha_{\rho}^d = \alpha_{\rho-1}^d$.
- 4: Metropolis step 2:
 Similarly for $\beta_{\rho-1}$ given θ_{ρ} and $\alpha_{\rho-1}$.
- 5: **end for**
- 6: Throw away all draws from the burn-in period $\rho = 1, \dots, B$.
- 7: **return** For all x and d , the estimated probabilities

$$\hat{p}^{dx} = \frac{1}{R} \sum_{\rho=B+1}^{B+R} \mathbf{1} \left(d = \arg \max_{d'} \theta_{\rho}^{d'x} \right). \quad (\text{A.1})$$

Denote by $\boldsymbol{\theta}, \mathbf{m}_t, \mathbf{r}_t$ the vectors of parameters, cumulative trials, and cumulative successes, where each of these is indexed by both d and x , and denote by $\boldsymbol{\alpha}, \boldsymbol{\beta}$ the vectors of hyperparameters indexed by d . Let ρ index replication draws, with ρ ranging from 1 to $B + R$. We sample from the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ given $\mathbf{m}_{t-1}, \mathbf{r}_{t-1}$ using the Markov Chain Monte Carlo algorithm described in Algorithm 1. Markov Chain Monte Carlo methods are reviewed in [Gelman et al. \(2014\)](#), chapter 11.

Algorithm 1 converges to a stationary distribution that equals the joint posterior of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\theta}$ given $\mathbf{m}_t, \mathbf{r}_t$. In particular, we have that the posterior probability that a treatment d is optimal given x , in the sense that it maximizes the probability of employment, is given by

$$p_t^{dx} = P \left(d = \arg \max_{d'} \theta^{d'x} | \mathbf{m}_t, \mathbf{r}_t \right) = \text{plim}_{R \rightarrow \infty} \frac{1}{R} \sum_{\rho=1}^R \mathbf{1} \left(d = \arg \max_{d'} \theta_{\rho}^{d'x} \right). \quad (\text{A.2})$$

In our implementation of this algorithm, we use a warm-up period of $B = 1,000$, and then draw $R = 10,000$ replications; averaging over these gives our estimated posterior distribution. These values are generously chosen relative to standard recommendations (cf. [Gelman et al. \(2014\)](#) chapter 11), making convergence likely. In our simulations these values yield stable posterior probabilities.

A.3 Additional tables and figures

Table A.1: Sampling methods by nationality

Sampling method	Jordanian		Syrian	
	Number	Percentage	Number	Percentage
Referral	662	31%	577	35%
Community-based Organization	753	36%	360	22%
Home visit	167	8%	420	25%
Social media	101	5%	29	2%
UNHCR visit	3	0%	95	6%
IRC office visit	405	19%	178	11%
Other	16	1%	4	0%
Total	2107		1663	

Table A.2: **Observations by stratum and treatment**

	Cash	Information	Nudge	Control
Syr, M, < HS, never emp	51	35	61	58
Syr, M, < HS, ever emp	86	75	102	152
Syr, M, >= HS, never emp	3	3	3	4
Syr, M, >= HS, ever emp	4	2	11	12
Syr, F, < HS, never emp	244	111	151	156
Syr, F, < HS, ever emp	61	32	89	89
Syr, F, >= HS, never emp	10	5	10	9
Syr, F, >= HS, ever emp	3	3	5	5
Jor, M, < HS, never emp	47	44	44	106
Jor, M, < HS, ever emp	40	90	120	110
Jor, M, >= HS, never emp	18	23	12	9
Jor, M, >= HS, ever emp	47	23	27	65
Jor, F, < HS, never emp	101	193	153	117
Jor, F, < HS, ever emp	65	68	78	54
Jor, F, >= HS, never emp	58	52	60	48
Jor, F, >= HS, ever emp	22	23	87	60

Table A.3: **Successes by stratum and treatment**

	Cash	Information	Nudge	Control
Syr, M, < HS, never emp	2	0	2	2
Syr, M, < HS, ever emp	6	3	6	9
Syr, M, >= HS, never emp	1	0	0	1
Syr, M, >= HS, ever emp	0	0	1	1
Syr, F, < HS, never emp	6	2	1	0
Syr, F, < HS, ever emp	1	1	3	2
Syr, F, >= HS, never emp	1	0	1	0
Syr, F, >= HS, ever emp	0	0	0	0
Jor, M, < HS, never emp	2	1	3	8
Jor, M, < HS, ever emp	4	9	13	13
Jor, M, >= HS, never emp	3	4	1	0
Jor, M, >= HS, ever emp	4	2	1	6
Jor, F, < HS, never emp	2	9	8	4
Jor, F, < HS, ever emp	9	8	4	4
Jor, F, >= HS, never emp	5	2	1	3
Jor, F, >= HS, ever emp	0	0	13	3

Note: The table reports results for wage employment at the time of the six weeks follow-up interview.

Table A.4: Success rates by stratum and treatment

	Cash	Information	Nudge	Control
Syr, M, < HS, never emp	0.04	0.00	0.03	0.03
Syr, M, < HS, ever emp	0.07	0.04	0.06	0.06
Syr, M, >= HS, never emp	0.33	0.00	0.00	0.25
Syr, M, >= HS, ever emp	0.00	0.00	0.09	0.08
Syr, F, < HS, never emp	0.02	0.02	0.01	0.00
Syr, F, < HS, ever emp	0.02	0.03	0.03	0.02
Syr, F, >= HS, never emp	0.10	0.00	0.10	0.00
Syr, F, >= HS, ever emp	0.00	0.00	0.00	0.00
Jor, M, < HS, never emp	0.04	0.02	0.07	0.08
Jor, M, < HS, ever emp	0.10	0.10	0.11	0.12
Jor, M, >= HS, never emp	0.17	0.17	0.08	0.00
Jor, M, >= HS, ever emp	0.09	0.09	0.04	0.09
Jor, F, < HS, never emp	0.02	0.05	0.05	0.03
Jor, F, < HS, ever emp	0.14	0.12	0.05	0.07
Jor, F, >= HS, never emp	0.09	0.04	0.02	0.06
Jor, F, >= HS, ever emp	0.00	0.00	0.15	0.05

Note: The table reports results for wage employment at the time of the six weeks follow-up interview.

Table A.5: **Weighted mean differences in employment by stratum, with randomisation inference p -values**

Subgroup	Treatment	Success rate	Δ	P-value
Female	Cash		0.010	0.211
Female	Information		0.005	0.342
Female	Nudge		0.011	0.201
Female	Control	0.031		
Male	Cash		-0.001	0.501
Male	Information		-0.020	0.857
Male	Nudge		-0.009	0.676
Male	Control	0.077		
Jordanian	Cash		-0.001	0.531
Jordanian	Information		-0.006	0.648
Jordanian	Nudge		0.002	0.463
Jordanian	Control	0.068		
Syrian	Cash		0.013	0.123
Syrian	Information		-0.004	0.626
Syrian	Nudge		0.005	0.348
Syrian	Control	0.027		
No high school	Cash		0.005	0.329
No high school	Information		-0.002	0.574
No high school	Nudge		0.002	0.428
No high school	Control	0.046		
High school	Cash		0.009	0.387
High school	Information		-0.015	0.723
High school	Nudge		0.007	0.405
High school	Control	0.061		
Never employed	Cash		0.011	0.206
Never employed	Information		-0.001	0.514
Never employed	Nudge		0.004	0.402
Never employed	Control	0.031		
Ever employed	Cash		0.000	0.501
Ever employed	Information		-0.010	0.730
Ever employed	Nudge		0.003	0.445
Ever employed	Control	0.071		

Note: The table reports results for wage employment at the time of the six weeks follow-up interview. Δ is the difference between weighted mean employment in a given treatment group and in the control group. p -values obtained with the randomization inference procedure discussed in Section 3.5.

Table A.6: 95% credible sets for average potential outcomes

stratum	Cash	Information	Nudge	Control
Syr, M, < HS, never emp	(0.010, 0.110)	(0.000, 0.080)	(0.010, 0.090)	(0.010, 0.100)
Syr, M, < HS, ever emp	(0.030, 0.120)	(0.010, 0.090)	(0.030, 0.100)	(0.030, 0.100)
Syr, M, >= HS, never emp	(0.020, 0.260)	(0.000, 0.170)	(0.010, 0.140)	(0.020, 0.240)
Syr, M, >= HS, ever emp	(0.010, 0.170)	(0.000, 0.170)	(0.020, 0.150)	(0.010, 0.180)
Syr, F, < HS, never emp	(0.010, 0.050)	(0.010, 0.060)	(0.000, 0.050)	(0.000, 0.030)
Syr, F, < HS, ever emp	(0.010, 0.080)	(0.010, 0.110)	(0.020, 0.080)	(0.010, 0.070)
Syr, F, >= HS, never emp	(0.020, 0.190)	(0.000, 0.150)	(0.020, 0.150)	(0.000, 0.140)
Syr, F, >= HS, ever emp	(0.010, 0.180)	(0.000, 0.160)	(0.010, 0.130)	(0.000, 0.160)
Jor, M, < HS, never emp	(0.010, 0.110)	(0.010, 0.090)	(0.020, 0.120)	(0.030, 0.120)
Jor, M, < HS, ever emp	(0.030, 0.170)	(0.040, 0.150)	(0.050, 0.140)	(0.060, 0.160)
Jor, M, >= HS, never emp	(0.040, 0.230)	(0.040, 0.220)	(0.020, 0.150)	(0.000, 0.140)
Jor, M, >= HS, ever emp	(0.030, 0.150)	(0.020, 0.160)	(0.010, 0.110)	(0.040, 0.150)
Jor, F, < HS, never emp	(0.010, 0.070)	(0.020, 0.080)	(0.030, 0.090)	(0.010, 0.080)
Jor, F, < HS, ever emp	(0.060, 0.190)	(0.050, 0.170)	(0.020, 0.100)	(0.030, 0.130)
Jor, F, >= HS, never emp	(0.030, 0.150)	(0.010, 0.100)	(0.010, 0.080)	(0.020, 0.130)
Jor, F, >= HS, ever emp	(0.000, 0.110)	(0.000, 0.100)	(0.060, 0.180)	(0.020, 0.110)

Note: The table reports results for wage employment at the time of the six weeks follow-up interview.

Table A.7: **Probability treatment is optimal, by stratum**

Stratum	Cash	Information	Nudge	Control
Syr, M, < HS, never emp	0.38	0.09	0.29	0.24
Syr, M, < HS, ever emp	0.44	0.10	0.23	0.23
Syr, M, >= HS, never emp	0.42	0.12	0.12	0.34
Syr, M, >= HS, ever emp	0.24	0.20	0.26	0.30
Syr, F, < HS, never emp	0.45	0.33	0.19	0.03
Syr, F, < HS, ever emp	0.19	0.35	0.33	0.13
Syr, F, >= HS, never emp	0.41	0.16	0.29	0.14
Syr, F, >= HS, ever emp	0.32	0.23	0.23	0.22
Jor, M, < HS, never emp	0.18	0.06	0.29	0.46
Jor, M, < HS, ever emp	0.20	0.18	0.20	0.41
Jor, M, >= HS, never emp	0.41	0.45	0.09	0.05
Jor, M, >= HS, ever emp	0.31	0.24	0.08	0.36
Jor, F, < HS, never emp	0.08	0.29	0.48	0.15
Jor, F, < HS, ever emp	0.58	0.32	0.02	0.09
Jor, F, >= HS, never emp	0.58	0.10	0.04	0.27
Jor, F, >= HS, ever emp	0.04	0.02	0.89	0.05

Table A.8: **Main outcomes**

Outcome	Definition
Wage employment	A dummy for whether the respondent currently has a wage-paying job.
Earnings	Earnings from main job (0 if not in wage employment).
Well-being	An index that comprises (i) monthly expenditure, (ii) life satisfaction (0-10 scale), (iii) an indicator of negative affect (feeling anxious on previous day on a 0-10 scale), (iv) an indicator of positive affect (feeling happy on previous day on a 0-10 scale).
Social integration	An index of seven social integration questions (each question asks the respondent to report on a scale from 1 to 5 how much he or she agrees with a given statement, for example, 'I feel connected to Jordan').
Intends to migrate	A dummy for whether the respondent intends to migrate to a different country in next 12 months (this does not include return migration to Syria).

Note: All indices are constructed using the method outlined in [Anderson \(2008\)](#).

Table A.9: **Treatment effects on employment for Syrians, after 2 months**

	Employed below 200 JOD	Employed above 200 JOD
Cash	0.033 (0.042)	0.019 (0.117)
Information	0.048 (0.007)	-0.002 (0.537)
Nudge	0.027 (0.077)	0.006 (0.354)
Control mean	0.043	0.048
Observations	1623	1608

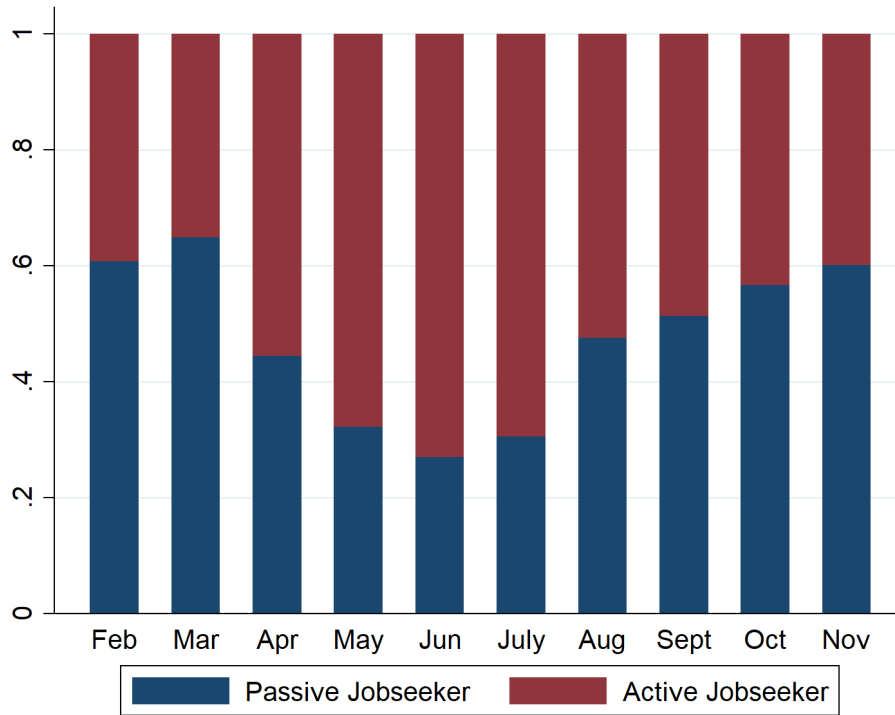
Table A.10: **Treatment effects after 2 months, full sample**

	Employed	Earnings	Well-being
Cash	0.003 (0.420)	-1.167 (0.626)	0.057 (0.036)
Information	0.003 (0.404)	-0.298 (0.537)	0.024 (0.217)
Nudge	0.000 (0.521)	-1.444 (0.680)	0.046 (0.075)
Control mean	0.103	22.758	0.025
Observations	3478	3463	3478

Table A.11: **Treatment effects after 4 months, full sample**

	Employed	Earnings	Well-being
Cash	0.011 (0.190)	0.974 (0.356)	0.064 (0.002)
Information	0.011 (0.189)	0.955 (0.379)	0.020 (0.190)
Nudge	0.008 (0.279)	0.236 (0.462)	0.028 (0.113)
Control mean	0.107	22.827	0.051
Observations	3593	3590	3593

Figure A.1: Active and passive sampling



Note: This Figure reports the proportion of jobseekers selected through active and passive sampling.

Figure A.2: Distribution of observations across strata

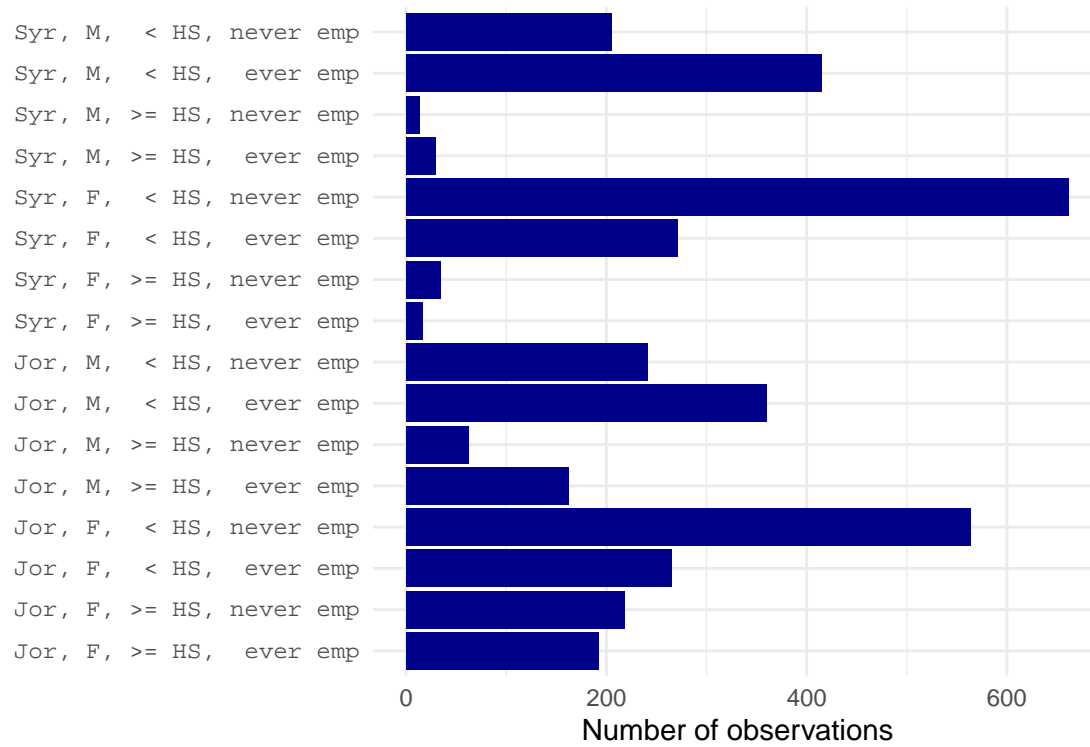


Figure A.3: Observations by week

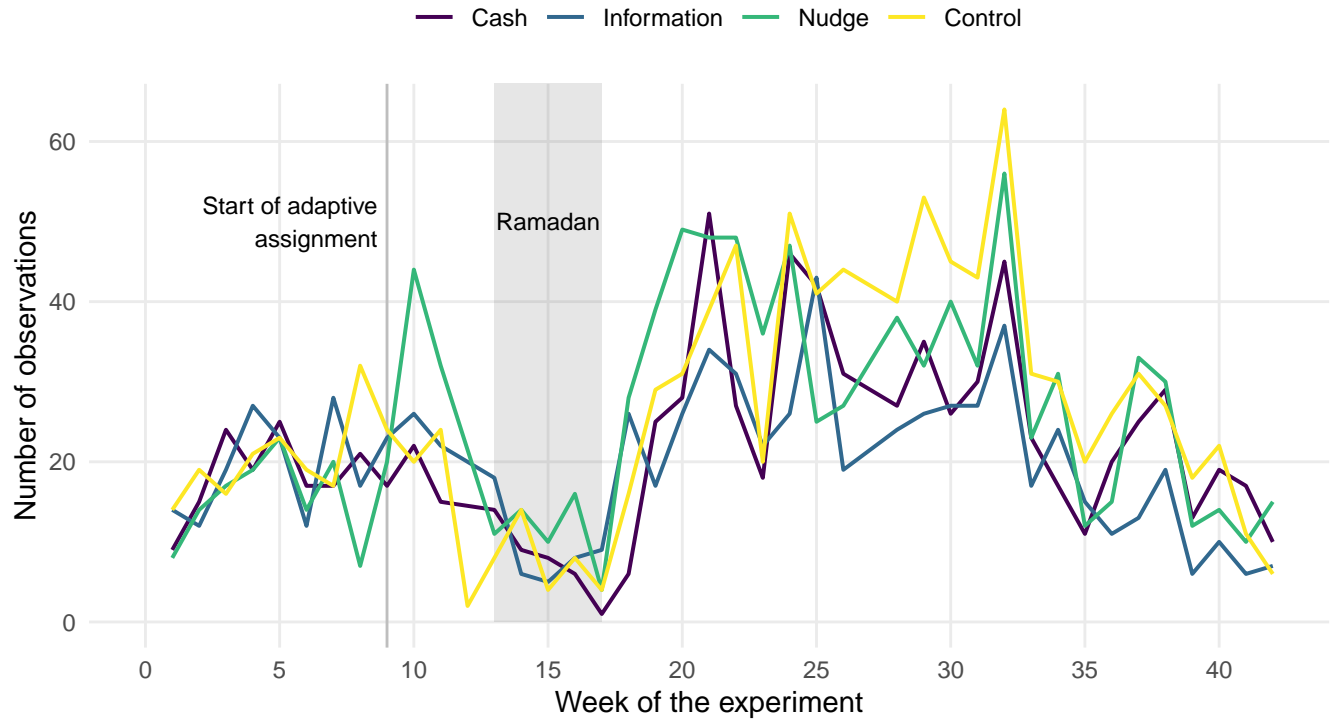


Figure A.4: Credible sets for average potential outcomes, and for average treatment effects relative to the control treatment

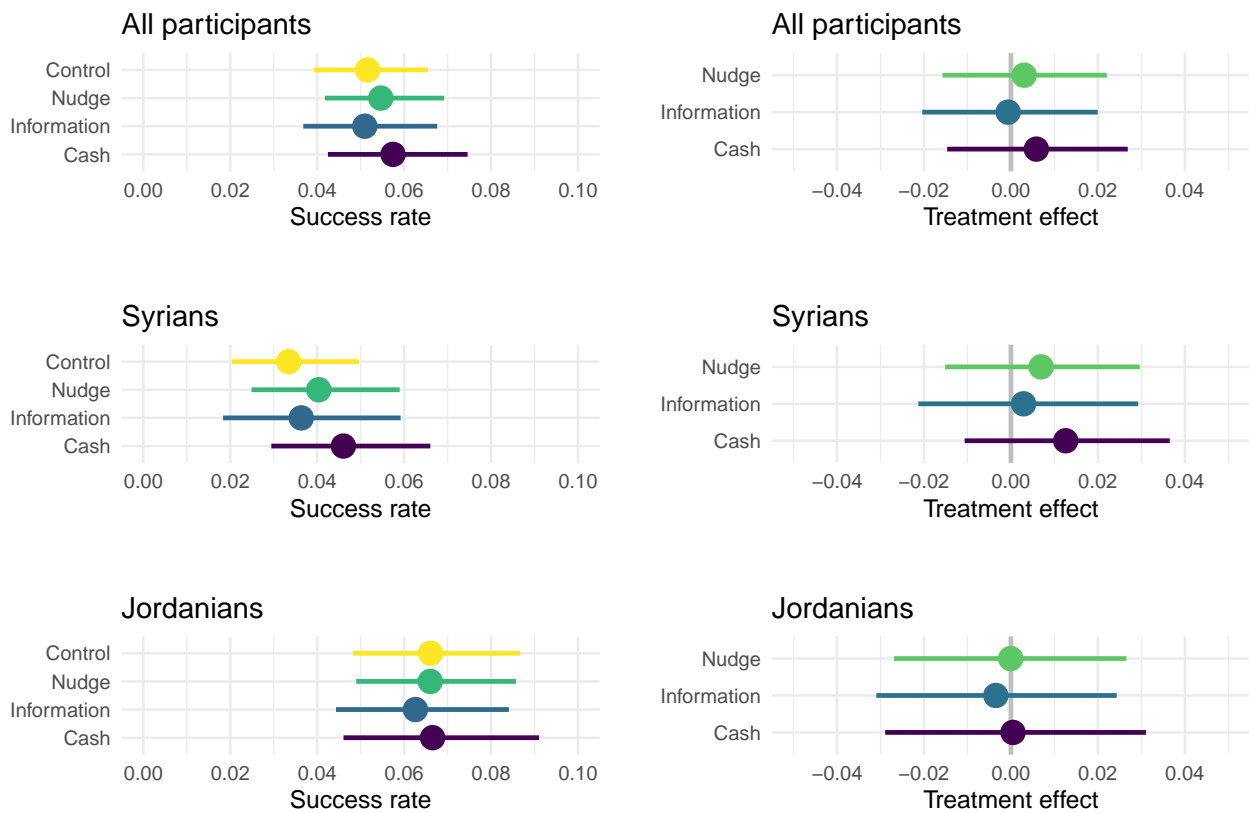


Figure A.5: 95% Credible sets for average potential outcomes across strata

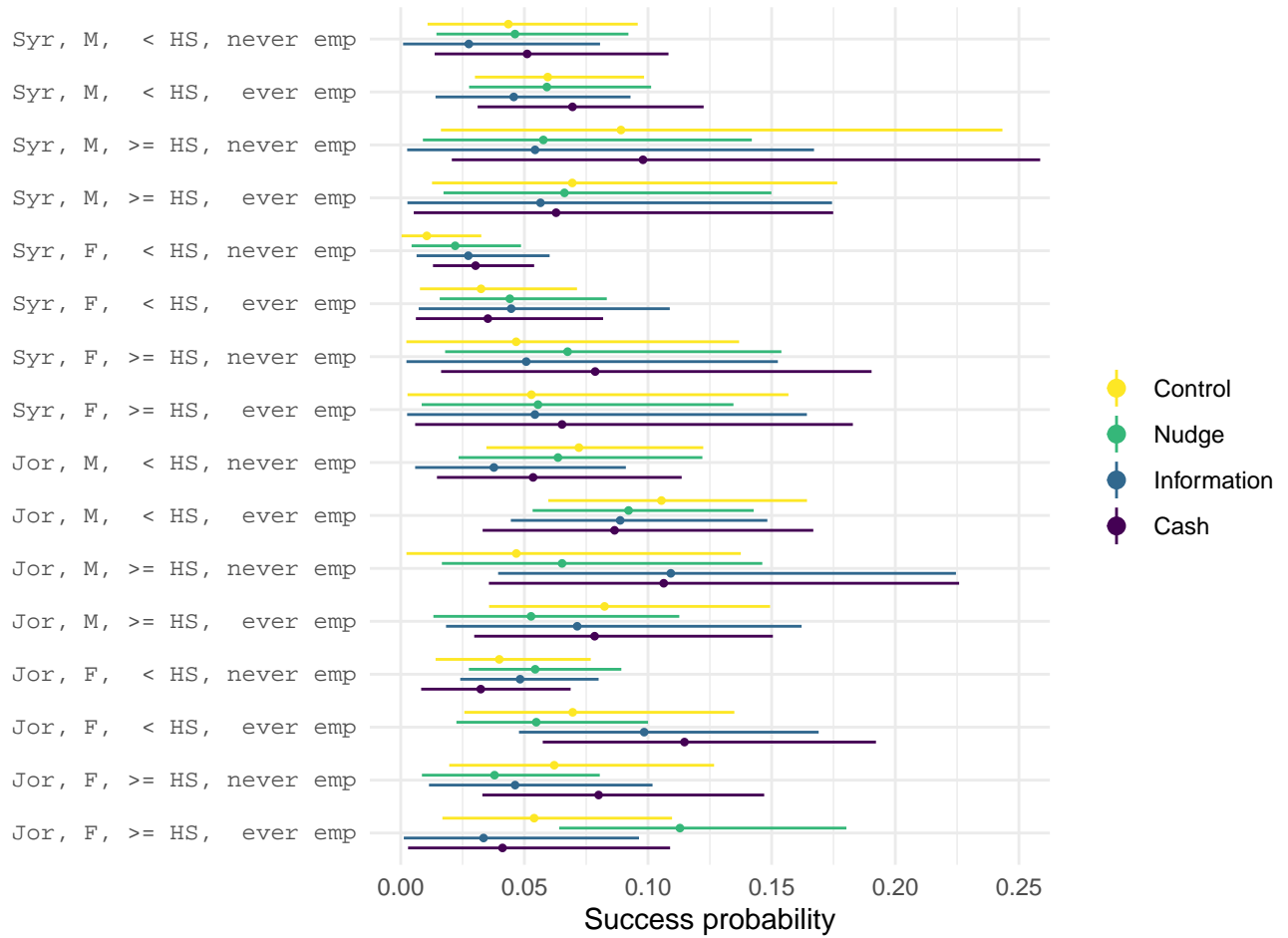


Figure A.6: Job search and baseline expenditure (control Syrians)

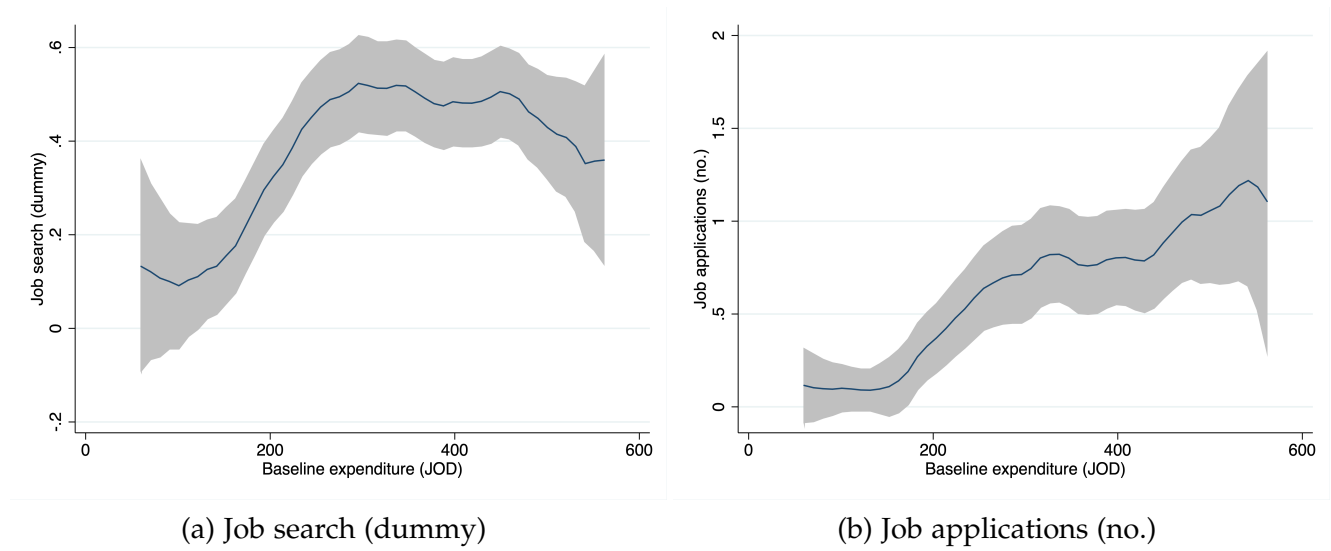


Figure A.7: Job search and baseline expenditure (control Jordanians)

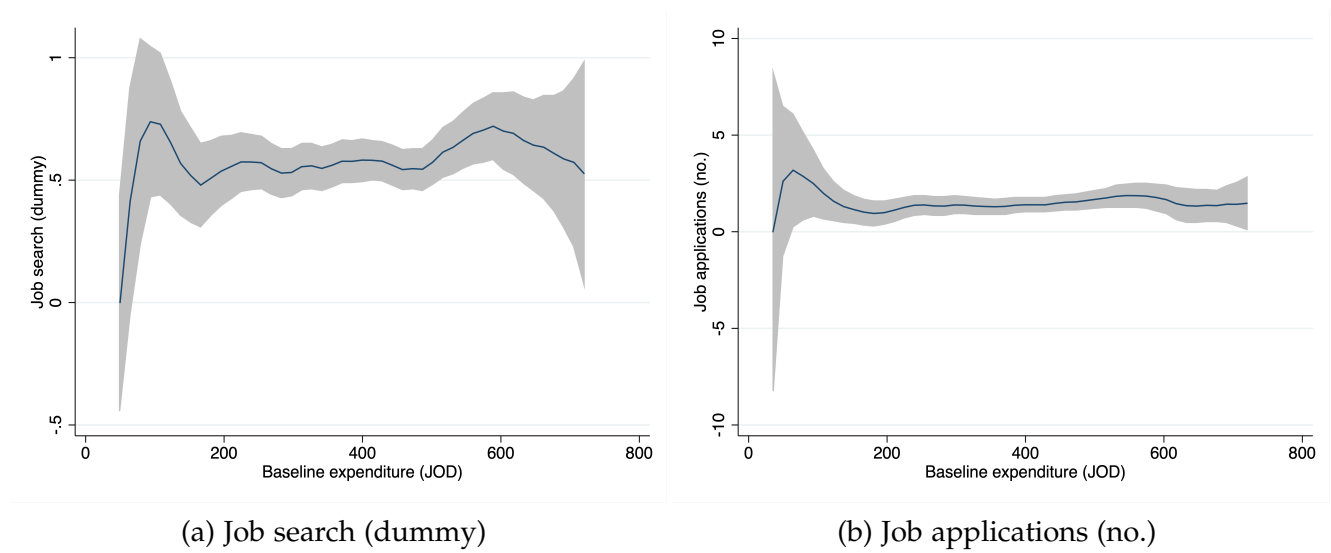


Figure A.8: Treatment effects on job search by baseline expenditure (Syrians)

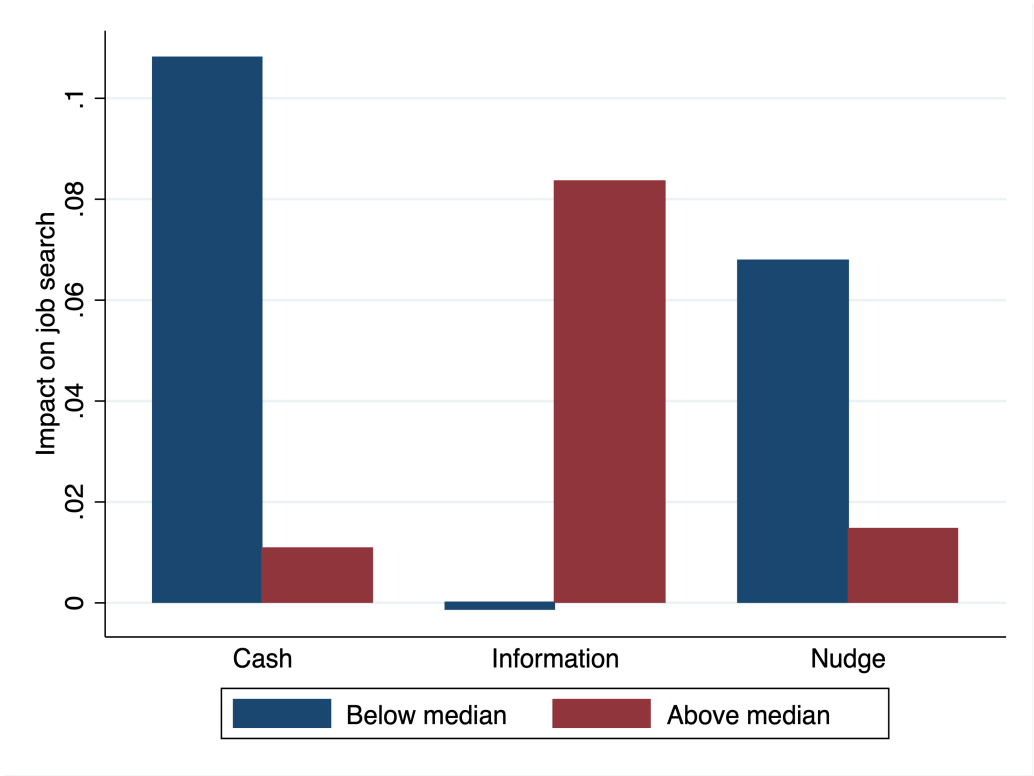


Figure A.9: Treatment effects on applications by baseline expenditure (Syrians)

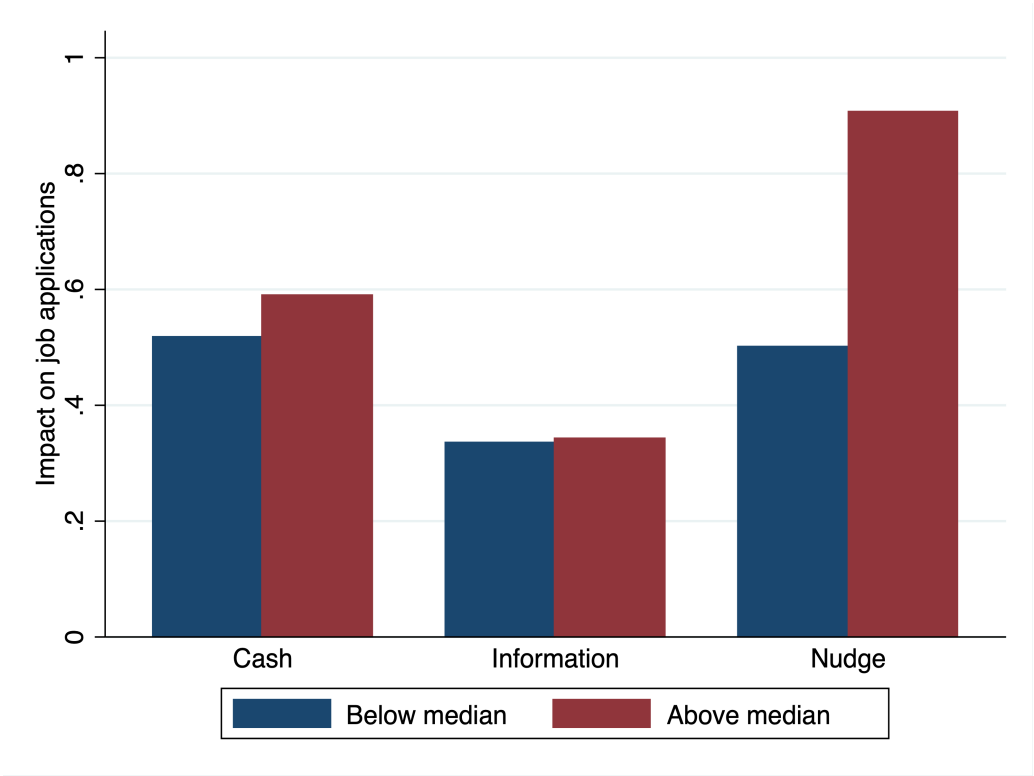


Figure A.10: Treatment effects on 4-month employment by baseline expenditure (Syrians)

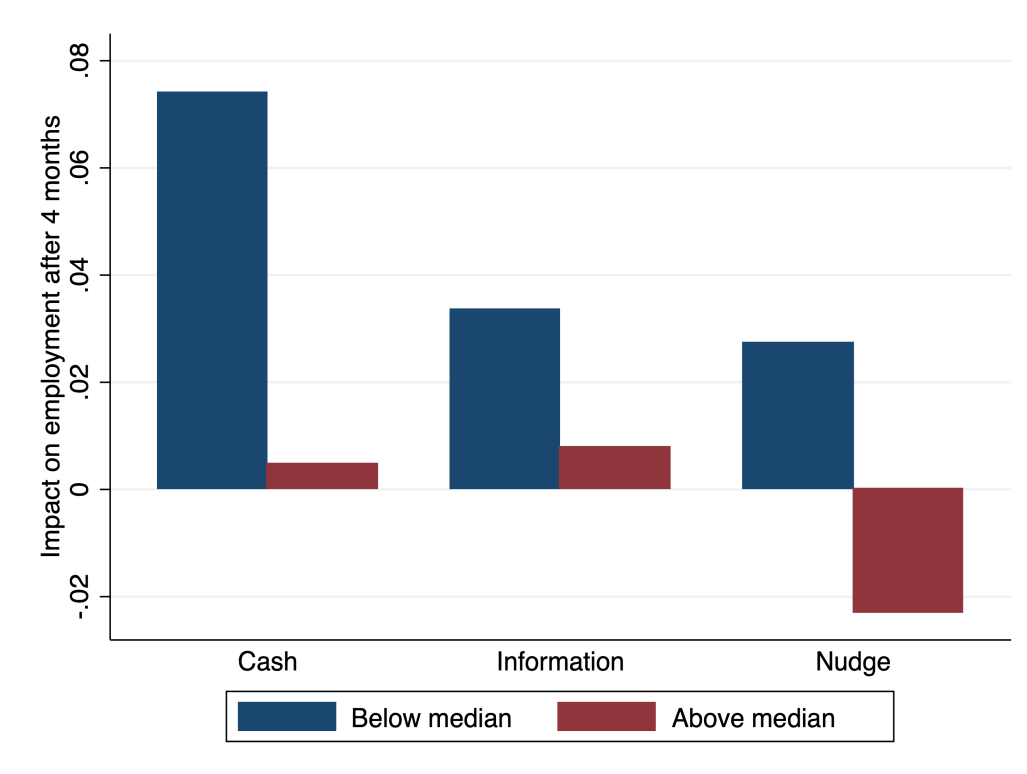
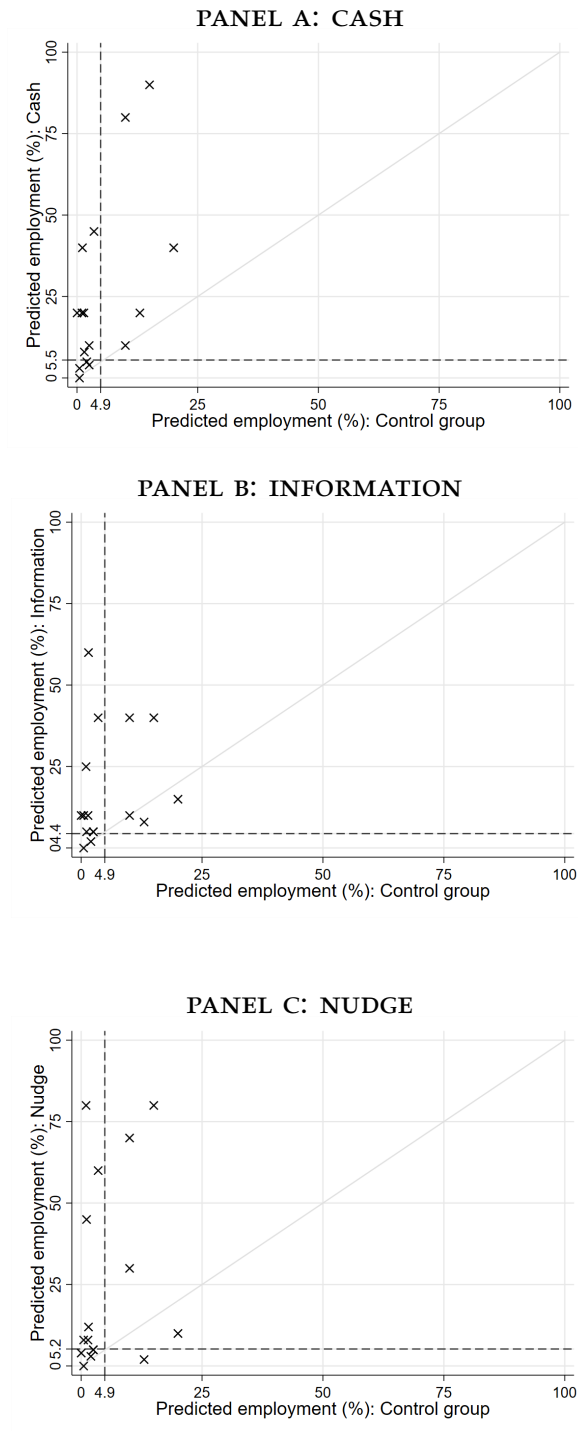
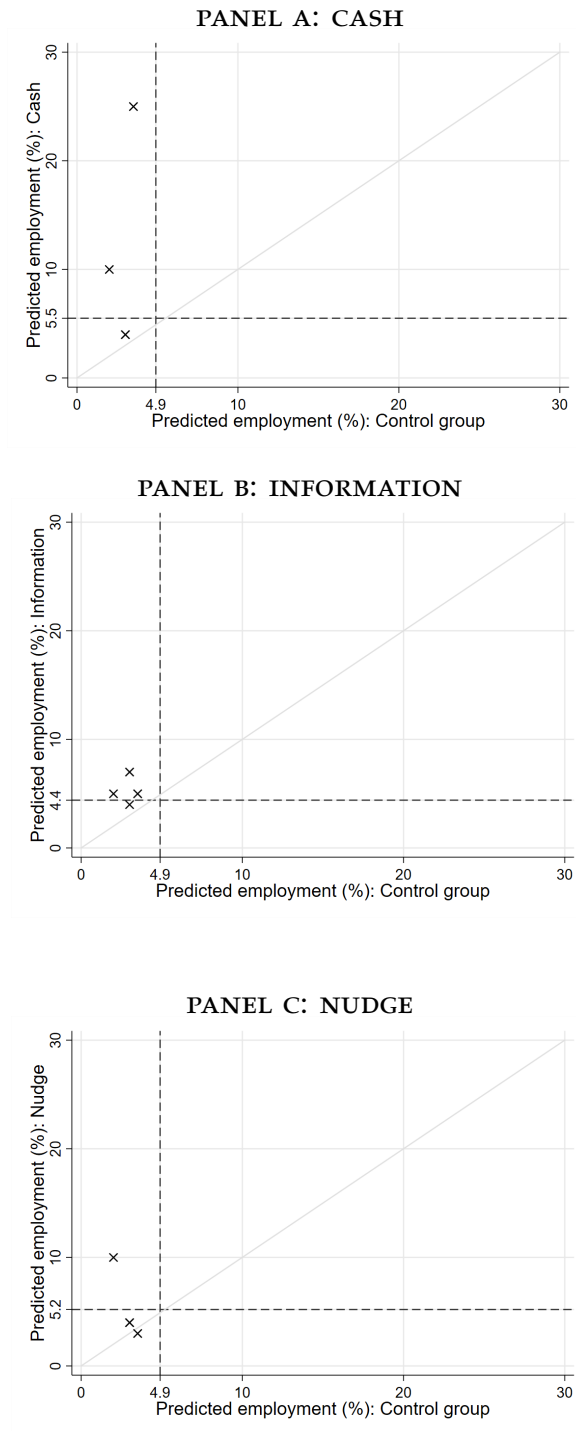


Figure A.11: Forecast employment outcomes: Local staff



Note: These scatterplots show IRC employees' incentivized forecasts of six-week employment rates under each of the three treatment arms; for each plot, we graph against the incentivized forecast of the six-week rate for the control group. On each plot, we superimpose the weighted average employment rates from Table A.5 in the paper.

Figure A.12: **Forecast employment outcomes: Head-office staff**

Note: These scatterplots show IRC employees' incentivized forecasts of six-week employment rates under each of the three treatment arms; for each plot, we graph against the incentivized forecast of the six-week rate for the control group. On each plot, we superimpose the weighted average employment rates from Table A.5 in the paper.